SYMMETRIC DECREASING REARRANGEMENT IS SOMETIMES CONTINUOUS

FREDERICK J. ALMGREN, JR. AND ELLIOTT H. LIEB

CONTENTS

1. The Problem and its History
 - 1.1. The setting of the problem
 - 1.2. Co-area regular and co-area irregular functions
 - 1.2.1. Definition. $\mathcal{E}_0$
 - 1.2.2. Definition. $\mathcal{E}_1$
 - 1.2.3. Definition. $\mathcal{E}_2$
 - 1.2.4. Notation. χ_A
 - 1.2.5. Definition. Distribution functions, co-area distribution functions, residual distribution functions
 - 1.2.6. Definition. Co-area regular functions, co-area irregular functions
 - 1.3. Definition. Symmetric decreasing rearrangement
 - 1.4. Theorem. Rearrangement is strongly continuous in $\mathbf{W}^{1,p}(\mathbf{R}^n)$
 - 1.5. A counterexample to a possible extension of Coron's $\mathbf{R}^1$ theorem
2. Suitable Integrands and Basic Properties of Rearrangement
 - 2.1. Definition. Suitable integrands
 - 2.2. Theorem. An extension of Riesz's rearrangement inequality
 - 2.3. Corollary. A rearrangement inequality for convex integrands
 - 2.4. Theorem. Convergence in measure
 - 2.5. Theorem. Functions in $\mathcal{E}_2$ belong to $\mathbf{L}^p$
 - 2.6. Theorem. Rearrangement preserves $\mathcal{E}_2$
 - 2.7. Theorem. Rearrangement decreases integrals of convex integrands
 - 2.8. Extension of symmetric decreasing rearrangement to vector-valued functions
3. Properties of the Co-area Distribution Function $\mathcal{F}_f$
 - 3.1. Theorem. Gradients vanish on the inverse image of small sets
 - 3.2. Theorem. $\mathcal{F}_f$ is absolutely continuous
 - 3.3. Theorem. Equi-absolute continuity
 - 3.4. Theorem. Convergence of cutoff $\mathcal{F}_j$'s
 - 3.5. Theorem. Upper semicontinuity of the co-area distribution functions $\mathcal{F}_j$
 - 3.6. The co-area formula and co-area regularity

Received by the editors February 21, 1989. The results of this paper were presented to a Conference on Geometry of Solutions to PDE, Cortona, June 1988 and to a Workshop on Variational Problems, Paris, June 1988 as summarized in [AL1]. They were also announced in [AL2].

1980 *Mathematics Subject Classification* (1985 Revision). Primary 46E35; Secondary 26B99, 47B38.

The first author was supported in part by NSF grant DMS-8701617-A01.

The second author was supported in part by NSF grant PHY-85-15288-A03.

4. The Main Theorems

- 4.1. Theorem. Derivatives of distribution functions and the continuity of rearrangement
- 4.2. Theorem. Rearrangement is continuous at co-area regular functions
- 4.3. Theorem. Rearrangement is not continuous at co-area irregular functions

5. Existence of Co-Area Irregular Functions

- 5.1. Construction of a $C^{n-1,\lambda}(\mathbf{R}^n)$ co-area irregular function
- 5.2. Theorem. Sufficient differentiability implies co-area regularity
- 5.3. Theorem. Co-area regular and co-area irregular functions are dense in $\mathcal{E}_1$
- 5.4. Theorem. If $n = 1$, there are no co-area irregular functions

6. Rearrangement is Not Continuous at Co-area Irregular Functions

- 6.1. Proof of Theorem 4.3
- 6.2. Theorem. Rearrangement need not be continuous even for one-dimensional affine perturbations

7. Convergence with General Integrands is Effectively Equivalent to $\mathbf{W}^{1,1}$ Convergence

- 7.1. Definition. I -slab functions
- 7.2. Theorem. Ψ convergence effectively implies $\mathbf{W}^{1,1}$ convergence
- 7.3. Theorem. General Ψ convergence of slab functions
- 7.4. Theorem. $\mathbf{W}^{1,1}$ convergence implies convergence for convex integrands and $\mathcal{E}_2$ functions
- 7.5. Theorem. $\mathbf{W}^{1,1}$ convergence implies convergence for convex integrands and $\mathcal{E}_1$ functions

8. Arc Length Convergence and $\mathbf{W}^{1,1}$ Convergence

- 8.1. Definition. Arc length over an interval
- 8.2. Proposition. Upper semicontinuity can imply continuity
- 8.3. Theorem. Arc length integrals and convergence of functions
- 8.4. Generalized graphs and inverses of monotone functions and their arc length integrals
- 8.5. Theorem. Equality of arc length integrals
- 8.6. Theorem. Convergence with strictly convex integrands implies $\mathbf{W}^{1,1}$ convergence
- 8.7. Theorem. Arc length convergence implies $\mathbf{W}^{1,1}(\mathbf{R})$ convergence
- 8.8. Proof of Theorem 4.1
- 8.9. Proof of Theorem 4.2

9. Symmetric Rearrangement is Strongly Continuous for Fractional Derivatives

- 9.1. Fractional Sobolev spaces
- 9.2. Theorem. Rearrangement is continuous in $\mathbf{W}^{\alpha,p}$

References

1. THE PROBLEM AND ITS HISTORY

1.1. The setting of the problem. Suppose $f(x_1, x_2) \geq 0$ is a continuously differentiable function supported in the unit disk in the plane. Its symmetric decreasing rearrangement is the rotationally invariant function $f^*(x_1, x_2)$ whose level sets are circles enclosing the same area as the level sets of f . Such rearrangement preserves $\mathbf{L}^p$ norms and generally decreases convex gradient integrals, e.g., $\|\nabla f^*\|_p \leq \|\nabla f\|_p$ ($1 \leq p < \infty$). Now suppose that $f_j(x_1, x_2) \geq 0$ ($j = 1, 2, 3, \dots$) is a sequence of infinitely differentiable functions also supported in the unit disk which converge uniformly together with first derivatives to f . The symmetrized functions also converge uniformly. The real question is about convergence of the derivatives of the symmetrized functions. One of the principal results of this paper is a proof of our discovery that *the derivatives*

of the symmetrized functions need not converge strongly, e.g., it can happen that $\|\nabla f_j^* - \nabla f^*\|_p \not\rightarrow 0$ for every p . We further characterize exactly those f 's for which convergence is assured and for which it can fail.

Associated with each real-valued function f belonging to the function space $\mathbf{W}^{1,p}(\mathbf{R}^n)$ is a symmetric decreasing rearrangement function $f^* \equiv \mathcal{R}(f)$. (Formal definitions are in 1.3 below.) The rearrangement map $\mathcal{R}: f \rightarrow f^*$ in general dimensions also decreases gradient norms. For this reason alone, rearrangement has long been a basic tool in the calculus of variations and in the theory of those partial differential equations (PDE's) that arise as Euler-Lagrange equations of variational problems; it permits one to concentrate attention on radial, monotone functions and thereby reduces many problems to simple one-dimensional ones. Some examples are (i) the lowest eigenfunction of the Laplacian in a ball is symmetric decreasing; (ii) the body with smallest capacity for a given volume is a ball [PS]; (iii) the optimal functions for the Sobolev and Hardy-Littlewood-Sobolev inequalities are symmetric decreasing and can be explicitly calculated [LE2]. Other examples are given in [BC, KB].

Obviously, $\mathcal{R}$ is highly nonlocal, nonlinear, and nonintuitive, but the property of decreasing gradient norms would lead one to surmise that $\mathcal{R}$ is a smoothing operator in some sense. Thus, when W. Ni and L. Nirenberg asked some years ago whether $\mathcal{R}$ is continuous in the $\mathbf{W}^{1,p}$ topology, the answer appeared to be that it should be so (it is easy to prove that $\mathcal{R}$ is always a contraction in $\mathbf{L}^p$). Indeed, by an elegant analysis, Coron [CJ] proved this in $\mathbf{R}^1$. An affirmative answer to this question would have meant that the mountain pass lemma could be used to establish spherically symmetric solutions of certain PDE's, and Coron's result led to just such an application [RS]. (R. E. L. Turner pointed out to us that sometimes convexity can substitute for continuity in applications of the mountain pass lemma [AAT, BBT].) Our result is that $\mathcal{R}$ is not continuous in $\mathbf{W}^{1,p}(\mathbf{R}^n)$ for $n \geq 2$, and it is surprising to us at least. Since almost all applications of $\mathcal{R}$ —apart from the mountain pass application—do not rely on continuity, our result does not have much immediate impact on applications. It reveals, however, an unexpected subtlety about the geometry of level sets of functions and shows that intuition can be very wrong.

It has been long known [BC, KB, BZ, PS, HK, AT, DG, SE1, SE2, TG, LE1] that

$$\int |\nabla f|^p d\mathcal{L}^n \geq \int |\nabla f^*|^p d\mathcal{L}^n;$$

here $\mathcal{L}^n$ denotes Lebesgue's n -dimensional measure over $\mathbf{R}^n$ (see Theorems 2.6 and 2.7). Suppose that $\{f_j\}_j$ is a sequence of functions belonging to $\mathbf{W}^{1,p}(\mathbf{R}^n)$ that converges in norm to a limit function f . The general question which we investigate and answer in this paper is this: *under what conditions is it true that the f_j^* 's converge to f^* in the $\mathbf{W}^{1,p}(\mathbf{R}^n)$ norm?* We cannot vouch for the origin of this question, but as indicated above, one of us (Elliott Lieb) first heard it posed some years ago by W. Ni and L. Nirenberg.

Let us explain heuristically why this question does not have an obvious answer. The $\mathbf{L}^p(\mathbf{R}^n)$ part of the norm causes no difficulties. In fact, for any two

functions f and g we have

$$\|f - g\|_{L^p} \geq \|f^* - g^*\|_{L^p}.$$

This inequality is well known [CG, CT] and will be proved here in a more general context in Corollary 2.3. Thus $\|f_j^* - f^*\|_{L^p} \rightarrow 0$ as $j \rightarrow \infty$ since $\|f_j - f\|_{L^p} \rightarrow 0$.

It is the $L^p(\mathbf{R}^n)$ norm of the gradient which is problematic. Despite the fact that rearrangement cannot increase the $L^p(\mathbf{R}^n)$ norm of each ∇f_j , there can be no inequality of the form

$$|||\nabla f - \nabla g|||_{L^p} \geq C |||\nabla f^* - \nabla g^*|||_{L^p}$$

for any positive number C . To appreciate this fact, take as an example $n = 1$ and let $f(x) = g(x)$ for $x < 0$ but $f(x) \neq g(x)$ for $x > 0$. Let ∇f be very large (in $L^p(\mathbf{R})$ norm) for $x < 0$, but both ∇f and ∇g are small for $x > 0$. Then $|||\nabla f - \nabla g|||_{L^p}$ is small because there is cancellation when $x < 0$. On the other hand, ∇f^* and ∇g^* can both be large (for $x < 0$ and $x > 0$) and, in general, there will be no cancellation because rearrangement shifts the level sets. Thus, we can make $|||\nabla f^* - \nabla g^*|||_{L^p}$ as large as we please relative to $|||\nabla f - \nabla g|||_{L^p}$.

The foregoing notwithstanding, it was long felt that rearrangement is continuous, i.e., $|||\nabla f - \nabla g|||_{L^p} \rightarrow 0$ implies $|||\nabla f^* - \nabla g^*|||_{L^p} \rightarrow 0$. Indeed, as we mentioned, Coron in [CJ] proved this when $n = 1$. We shall comment further on Coron's result in 1.4 below, but for now let us record only that Coron's proof was very much limited to the case $n = 1$. As we discovered in an effort to extend his results, *the conjecture is false for $n \geq 2$* . We have, however, elucidated the precise condition under which continuity holds for $n \geq 2$; this requires the introduction of a new regularity concept for functions which we turn to next. However, as noted in 1.4 and proved in §9, the corresponding conjecture for the fractional Sobolev spaces $\mathbf{W}^{\alpha,p}(\mathbf{R}^n)$, $\alpha < 1$, is correct for all n . We thank Michael Loss for valuable discussions, especially about the $\mathbf{W}^{\alpha,p}(\mathbf{R}^n)$ case.

1.2. Co-area regular and co-area irregular functions. We first introduce the largest space of nonnegative functions for which it is possible to define rearrangements, namely,

1.2.1. Definition. We set $\mathbf{R}^+ = \{x : 0 \leq x < \infty\}$ and denote by $\mathcal{E}_0$ the space of $\mathcal{L}^n$ measurable functions f mapping $\mathbf{R}^n$ to $\mathbf{R}^+$ with the property that

$$\mathcal{L}^n(\{x : f(x) > y\}) < \infty \quad \text{for each positive number } y.$$

Within $\mathcal{E}_0$ is the set of functions needed to discuss gradients, namely,

1.2.2. Definition. We denote by $\mathcal{E}_1$ the subset of $\mathcal{E}_0$ consisting of functions f whose distribution gradients ∇f are functions belonging to $L^1_{\text{loc}}(\mathbf{R}^n, \mathbf{R}^n)$ and also for which the real-valued functions $|\nabla f|$ belong to $L^1(\Omega)$ for every $\mathcal{L}^n$ measurable set Ω of finite $\mathcal{L}^n$ measure.

Clearly, all nonnegative functions in $\mathbf{W}^{1,p}(\mathbf{R}^n)$ belong to $\mathcal{E}_1$ for all $p \geq 1$, because $\int_{\Omega} |\nabla f| d\mathcal{L}^n \leq \mathcal{L}^n(\Omega)^{1/p'} \|\nabla f\|_{L^p}$.

A still smaller set which we shall want to consider and which contains all non-negative functions in $\mathbf{W}^{1,1}(\mathbf{R}^n)$ but does *not* contain all nonnegative functions in $\mathbf{W}^{1,p}(\mathbf{R}^n)$ for $p > 1$ is given in the following definition.

1.2.3. Definition. We denote by $\mathcal{E}_2$ the subset of $\mathcal{E}_1$ consisting of functions f belonging to $\mathcal{E}_1$ for which $|\nabla f|$ belongs to $\mathbf{L}^1(\mathbf{R}^n)$.

1.2.4. Notation. It is convenient to use the notation $\chi_{\{A\}}$ symbolically to denote a function which takes value 1 when the test A is passed and takes value 0 otherwise; e.g., $\chi_{\{f>y\}}(x)$ equals 1 in case $f(x) > y$ and equals 0 otherwise. Generally, such a function $\chi_{\{A\}}$ will be defined only $\mathcal{L}^n$ almost everywhere, which is adequate for our purposes. The symbol $|\nabla f(x)|$ denotes the Euclidean norm of the vector $\nabla f(x)$. We shall say that a vector-valued function F (such as ∇f) belongs to $\mathbf{L}^p(\mathbf{R}^n)$ provided F is measurable and $|F|$ is in $\mathbf{L}^p(\mathbf{R}^n)$.

1.2.5. Definition. Associated with each function f belonging to $\mathcal{E}_1$ is its *distribution function* $\mathcal{V}_f: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ defined by setting

$$\mathcal{V}_f(y) = \int \chi_{\{f>y\}} d\mathcal{L}^n$$

for each $y > 0$. It is necessary for our purposes to write $\mathcal{V}_f$ as the sum of two other functions. The first of these is the *co-area distribution function* $\mathcal{F}_f$ which is defined by setting

$$\mathcal{F}_f(y) = \int \chi_{\{f>y\}} \chi_{\{\nabla f \neq 0\}} d\mathcal{L}^n$$

for each $y > 0$. The second summand we call the *residual distribution function* $\mathcal{G}_f$, and it is defined by setting

$$\mathcal{G}_f(y) = \int \chi_{\{f>y\}} \chi_{\{\nabla f = 0\}} d\mathcal{L}^n$$

for each $y > 0$.

The term “co-area distribution function” was suggested by H. Federer’s co-area formula [FH, 3.2.11], a mild extension of which shows that the function $\mathcal{F}_f$ is absolutely continuous on the interval (ε, ∞) for each positive number ε ; the formula furthermore readily exhibits the distribution first derivatives of $\mathcal{F}_f$ as integrals over level sets. In §3, we give an alternative proof of this absolute continuity.

Since each of the functions $\mathcal{V}_f$, $\mathcal{F}_f$, $\mathcal{G}_f$ is nonincreasing, the distribution derivatives $\mathcal{V}_f'$, $\mathcal{F}_f'$, $\mathcal{G}_f'$ are (negative) Radon measures on $(0, \infty)$. This means, in particular, that $\mathcal{V}_f'(K) < \infty$ for each compact subset of $(0, \infty)$ —it does not mean that $\mathcal{V}_f'(0, 1) < \infty$ which is generally not true. As indicated, the function $\mathcal{F}_f$ is absolutely continuous on each (ε, ∞) so that $\mathcal{F}_f'$ is absolutely continuous with respect to Lebesgue’s measure $\mathcal{L}^1$ on $(0, \infty)$. In order to avoid pedantry, we shall use the symbol $\mathcal{F}_f'$ both for the measure and for the

function which defines the measure, i.e., its derivative with respect to Lebesgue measure. The function $\mathcal{G}_f$, on the other hand, may or may not be absolutely continuous. In any case, the Radon measure $\mathcal{G}'_f$ on $(0, \infty)$ can be uniquely represented as the sum

$$\mathcal{G}'_f = -\mathcal{L}^1 \wedge h - \nu,$$

where $h: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is $\mathcal{L}^1$ summable on each (ε, ∞) and ν is singular with respect to $\mathcal{L}^1$. We can now make our fundamental definition.

1.2.6. Definition (Co-area regularity). A function f belonging to the space $\mathcal{E}_1$ is called *co-area regular* if and only if the Radon measure $\mathcal{G}'_f$ above is singular with respect to $\mathcal{L}^1$ on $(0, \infty)$. Otherwise, the function f is called *co-area irregular*.

If one thinks of the set on which $\nabla f = 0$, one would usually think of “flat spots,” or plateaus on which f is constant. These, however, contribute only to the singular part of our distribution derivative measure $\mathcal{G}'_f$. The absolutely continuous part of $\mathcal{G}'_f$ is more mysterious and harder to visualize. Somehow the $\nabla f = 0$ set is a set of positive $\mathcal{L}^n$ measure which is smeared out over uncountably many y values without there necessarily being any plateaus. Such functions exist when $n \geq 2$ (but not when $n = 1$, which is why Coron’s theorem is valid), and we construct an example of such a function f in 5.1. In fact, in our example, the distribution derivative measure $\mathcal{G}'_f$ is *purely absolutely continuous* with respect to $\mathcal{L}^1$. On the other hand, one can ask for a sufficient condition that f be co-area regular. One such condition is that f belong to $C_{\text{loc}}^{n-1,1}(\mathbf{R}^n)$, i.e., that f be $n-1$ times continuously differentiable and have $(n-1)$ st derivatives which satisfy a Lipschitz condition. This is essentially a special case of the general Morse-Sard-Federer theorem and is stated explicitly in Theorem 5.2 below. When $n = 1$, all functions in $\mathcal{E}_1$ are co-area regular, as Theorem 5.4 shows. When $n \geq 2$, the co-area regular functions and the co-area irregular functions each are dense in most interesting function spaces, as Theorem 5.3 shows.

1.3. Definition (Symmetric decreasing rearrangement). There are several possible definitions of $f^*: \mathbf{R}^n \rightarrow \mathbf{R}^+$, the *symmetric decreasing rearrangement* (rearrangement, in brief) of a function $f \in \mathcal{E}_0$. These various f^* ’s are equivalent in the sense that they agree up to sets of zero $\mathcal{L}^n$ measure. Our definition will be the following. Let $\alpha(n)$ denote the volume of a ball of radius one in $\mathbf{R}^n$ and for each $y > 0$ define the radius $R(y) = R_f(y)$ by

$$\alpha(n)R(y)^n = \int \chi_{\{f>y\}} d\mathcal{L}^n = \mathcal{V}_f(y).$$

Clearly, $R(y) < \infty$ since $f \in \mathcal{E}_0$. Let χ_R denote the characteristic function of the open ball of radius R centered at $0 \in \mathbf{R}^n$. Then

$$f^*(x) \equiv \int_{0 < y < \infty} \chi_{R(y)}(x) d\mathcal{L}^1 y.$$

In view of the fact that

$$f(x) = \int_{0 < y < \infty} \chi_{\{f > y\}}(x) d\mathcal{L}^1 y,$$

our definition amounts to replacing $\chi_{\{f > y\}}$ by $\chi_{R(y)}$, where

$$\int \chi_{\{f > y\}} d\mathcal{L}^n = \int \chi_{R(y)} d\mathcal{L}^n.$$

An equivalent definition is

$$f^*(x) = \sup\{y : R(y) > |x|\},$$

from which it is clear that f^* is lower semicontinuous. It is also clear that $\mathcal{L}^n(\{x : f(x) > y\}) = \mathcal{L}^n(\{x : f^*(x) > y\})$ for all $y > 0$.

With the aim of simplicity, we shall sometimes abuse notation by writing $f^*(|x|)$ (which is a map from $\mathbf{R}^+$ to $\mathbf{R}^+$) for $f^*(x)$.

Most of the time we will be considering functions from $\mathbf{R}^n$ to $\mathbf{R}^+$ in this paper. The main exception occurs in §9 in which we deal directly with functions from $\mathbf{R}^n$ to $\mathbf{C}$ simply because it was natural to do so in the context of our proof for $\mathbf{W}^{\alpha,p}(\mathbf{R}^n)$. Both in this case and in the general case, we would like to study symmetrizations of functions from $\mathbf{R}^n$ to $\mathbf{C}$ or from $\mathbf{R}^n$ to $\mathbf{R}^m$. In all such cases, one defines f^* to be $|f|^*$, where $|f(x)|$ is the Euclidean norm of $f(x)$. *The principal theorems in this paper extend to maps from $\mathbf{R}^n$ to $\mathbf{R}^m$ (or to $\mathbf{C}$).* There is a slight subtlety in carrying this out, which we discuss in 2.8.

With this preparation, we can now state our main theorem in a restricted context, namely, in $\mathbf{W}^{1,p}(\mathbf{R}^n)$. Later, in Theorems 4.1, 4.2, and 4.3, we shall state our results in full generality after we have introduced more general convex integrands than simply $|\xi|^p$.

1.4. Theorem (Rearrangement is strongly continuous in $\mathbf{W}^{1,p}(\mathbf{R}^n)$ only at co-area regular functions). *Let $1 \leq p < \infty$ and let $f, f_1, f_2, f_3, \dots$ be functions in $\mathbf{W}^{1,p}(\mathbf{R}^n)$ with $f_j \rightarrow f$ strongly as $j \rightarrow \infty$. If f is co-area regular, then $f_j^* \rightarrow f^*$ strongly in $\mathbf{W}^{1,p}(\mathbf{R}^n)$. On the other hand, if f belongs to $\mathbf{W}^{1,p}(\mathbf{R}^n)$ and is co-area irregular, then there exists some sequence $f_1, f_2, f_3, \dots$ of functions in $\mathbf{W}^{1,p}(\mathbf{R}^n)$ such that $f_j \rightarrow f$ strongly as $j \rightarrow \infty$ but $f_j^* \not\rightarrow f^*$ strongly in $\mathbf{W}^{1,p}(\mathbf{R}^n)$. Note in these cases that the criterion of co-area irregularity applies only to the limit function f .*

The proof of the second assertion of Theorem 1.4 about failure of convergence is given in §6. In §7, we show that strong convergence of general convex integrals is effectively equivalent to $\mathbf{W}^{1,1}$ convergence. In §8, we show that such $\mathbf{W}^{1,1}$ convergence is effectively equivalent to convergence of the arc lengths of the graphs of the functions $r \rightarrow f_j^*(r, \theta)$ to the arc lengths of the graph of the function $r \rightarrow f^*(r, \theta)$; here we are writing f^* in polar coordinates (r, θ) . In §8, we also show that co-area regularity of f implies the necessary arc length convergence.

It may seem odd that a theorem about convergence in $\mathbf{W}^{1,p}(\mathbf{R}^n)$ for $p > 1$ can be reduced to the $p = 1$ case. The underlying reason for this is the inequality (Theorem 2.7)

$$\|\nabla f\|_{\mathbf{L}^p} \geq \|\nabla f^*\|_{\mathbf{L}^p},$$

which imposes an a priori bound on the $\mathbf{L}^p(\mathbf{R}^n)$ norms of the ∇f_j^* 's in small horizontal slabs, given that the ∇f_j 's are $\mathbf{L}^p(\mathbf{R}^n)$ bounded in corresponding slabs.

A somewhat surprising fact, given the dichotomy presented in Theorem 1.4, is that one *always* has convergence of f_j^* to f^* for *fractional derivatives*, by which we mean the Sobolev spaces $\mathbf{W}^{\alpha,p}(\mathbf{R}^n)$ for $0 < \alpha < 1$ and $1 \leq p < \infty$. Co-area regularity plays no role here. This is proved in §9.

1.5. A counterexample to a possible extension of Coron's $\mathbf{R}^1$ theorem. For $n = 1$, Coron proved a stronger result for $\mathbf{W}^{1,p}(\mathbf{R})$ than $f_j^* \rightarrow f^*$ strongly when $f_j \rightarrow f$ strongly. He showed that the functional

$$\varphi(f) = \int |\nabla f|^p d\mathcal{L}^1 - C_p \int |\nabla f^*|^p d\mathcal{L}^1$$

is weakly lower semicontinuous when $C_p \leq 2^{-p}$. In view of Theorem 1.4, such an estimate cannot hold in $\mathbf{R}^n$ when $n \geq 2$ for any C_p because the map $f \rightarrow f^*$ is not even continuous. Before we fully appreciated this latter fact, we thought of imitating Coron's weak lower semicontinuity result and realized that the extension might not be valid when $n \geq 2$ and $p > 1$ for any C_p , for a reason totally unconnected with co-area irregularity. The difficulty, which we feel is worth noting here, is connected with a large distance phenomenon of *magnification of gradients under rearrangement* whereas co-area irregularity has to do with short distance behavior of the limit f . Thus, even within the category of co-area regular functions, for which strong convergence is assured, a Coron-type lower semicontinuity result cannot hold when $p > 1$.

Let us illustrate this magnification with an example. Let $f = F + G$, where F and G are functions which are spherically symmetric decreasing with respect to centers u and v in $\mathbf{R}^n$. The supports of F and G are disjoint. The large piece F is smooth, but it is close to being the characteristic function of a ball of radius $R \gg 1$, i.e., there is an $\varepsilon \ll 1$ such that when $R - \varepsilon < |x - u| < R$, then $|\nabla F|$ is very large; otherwise, $|\nabla F| = 0$. On the other hand, the support of G is contained in a ball of radius 1 and $|\nabla G|$ is approximately equal to 1 inside this ball. For example, $G(x) = \cos(\frac{\pi}{2}|x - v|)$ when $|x - v| \leq 1$. In particular, $f = F + G$ is co-area regular.

Now suppose f_j is a sequence in $\mathbf{W}^{1,p}(\mathbf{R}^n)$ which converges *weakly* to f , and $f_j = F + G_j$ with G_j spherically symmetric decreasing and with the support of G_j in the same ball $|x - v| \leq 1$. The assertion that φ is weakly lower

semicontinuous means that the drop in the first term, namely,

$$\begin{aligned} D_1 &= \lim_{j \rightarrow \infty} \inf \int |\nabla f_j|^p d\mathcal{L}^n - \int |\nabla f|^p d\mathcal{L}^n \\ &= \lim_{j \rightarrow \infty} \inf \int |\nabla G_j|^p d\mathcal{L}^n - \int |\nabla G|^p d\mathcal{L}^n \end{aligned}$$

is larger than C_p times the drop in the second term

$$D_2 = \lim_{j \rightarrow \infty} \inf \int |\nabla f_j^*|^p d\mathcal{L}^n - \int |\nabla f^*|^p d\mathcal{L}^n.$$

Note that D_1 involves the radial derivatives of G_j and G (but not F) multiplied by the volume factor r^{n-1} which is of order unity. On the other hand, f^* and the f_j^* 's are symmetric decreasing functions which are constant for $|x| < R - \varepsilon$, and then they decrease monotonically to zero in the region $R - \varepsilon < |x| < R + 1$. The radial derivatives of f^* (resp. f_j^*) are roughly $R^{n-1}/1^{n-1}$ times those of G (resp. G_j). The drop D_2 is then roughly the drop D_1 , but multiplied by the ratio $(R^{n-1}/1^{n-1})^p$. Since R can be made arbitrarily large, we see that we can make $CD_2 > D_1$ for any fixed choice of C .

2. SUITABLE INTEGRANDS AND BASIC PROPERTIES OF REARRANGEMENT

Our goal in this paper is to relate the convergence $\nabla f_j \rightarrow \nabla f$ of gradients of functions to the convergence $\nabla f_j^* \rightarrow \nabla f^*$ of the gradients of their symmetrizations. The notions of convergence we consider are given by integrals $\int \Psi(|\nabla f|) d\mathcal{L}^n$ and $\int \Psi(|\nabla f_j - \nabla f|) d\mathcal{L}^n$ associated with suitable integrands $\Psi: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ in accordance with the following definition.

2.1. Definition (Suitable integrands). By a *suitable integrand* we mean a convex function $\Psi: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ (not identically zero) with $\Psi(0) = 0$ and for which there is some finite constant $C \geq 1$ such that

$$\Psi(2t) \leq C\Psi(t) \quad \text{for each } t > 0.$$

This later requirement is the so-called Δ_2 condition in Orlicz space theory. Note that each suitable integrand is continuous and strictly increasing.

The following Theorem 2.2 was proved some time ago by H. Berestycki and E. Lieb, but it appears here for the first time. It is a generalization of Theorem 3 of [CRZ] (in fact, they explicitly refer to this theorem on p. 433); it is also a generalization of the basic Riesz rearrangement inequality [RF]; see also [HLP]. We will use it in this part to resolve questions about the convergence $f_j^* \rightarrow f^*$ of symmetrized functions. It will be needed again in the proofs in §9.

2.2. Theorem (An extension of Riesz's rearrangement inequality). Let $F: \mathbf{R}^+ \times \mathbf{R}^+ \rightarrow \mathbf{R}^+$ be a continuous function such that $F(0, 0) = 0$ and

$$(1) \quad F(u_2, v_2) + F(u_1, v_1) \geq F(u_2, v_1) + F(u_1, v_2)$$

whenever $u_2 \geq u_1 \geq 0$ and $v_2 \geq v_1 \geq 0$. (For twice continuously differentiable functions, (1) is equivalent to the assumption that $F_{12} \equiv (\partial^2 F / \partial u \partial v) \geq 0$.) Then whenever f and g are functions belonging to our class $\mathcal{C}_0$, we have the inequalities

$$(2) \quad \begin{aligned} & \iint F(f(x), g(y)) W(ax + by) d\mathcal{L}^n x d\mathcal{L}^n y \\ & \leq \iint F(f^*(x), g^*(x)) W^*(ax + by) d\mathcal{L}^n x d\mathcal{L}^n y. \end{aligned}$$

and

$$\int F(f(x), g(x)) d\mathcal{L}^n x \leq \int F(f^*(x), g^*(x)) d\mathcal{L}^n x$$

for every nonnegative function W in $L^1(\mathbf{R}^n)$ and every choice of nonzero numbers a and b .

Proof. We first prove (2). We assume that the right-hand side of (2) is finite (since, otherwise, there is nothing to prove). Because of Fubini's theorem, Euclidean invariance, scaling of Lebesgue's measure, and the fact that $[f(cx)]^* = f^*(cx)$, it is sufficient to prove the theorem under the additional assumptions that $a = 1$ and $b = -1$. From the definition of rearrangement, it follows that

$$\int F(f(x), 0) d\mathcal{L}^n = \int F(f^*(x), 0) d\mathcal{L}^n,$$

etc. The Euclidean invariance of Lebesgue's measure implies that the integral $\int W(x - y) d\mathcal{L}^n y$ is independent of x , etc. We combine these observations with the inequality $F(u, 0) + F(0, v) \leq F(u, v)$ to conclude the equality

$$\begin{aligned} & \iint [F(f(x), 0) + F(0, g(y))] W(x - y) d\mathcal{L}^n x d\mathcal{L}^n y \\ & = \iint [F(f^*(x), 0) + F(0, g^*(x))] W^*(x - y) d\mathcal{L}^n x d\mathcal{L}^n y < \infty. \end{aligned}$$

Consider the function $\tilde{F}(u, v) = F(u, v) - F(u, 0) - F(v, 0)$. Inequality (1) continues to hold if we replace F by $\tilde{F}$. Furthermore, in view of our last integral equality, inequality (2) holds if and only if the corresponding inequality with $\tilde{F}$ replacing F holds. We note that the function $\tilde{F}$ satisfies the condition $\tilde{F}(0, v) = \tilde{F}(u, 0) = 0$. It is therefore sufficient to prove the theorem under the additional assumption that $F(0, v) = F(u, 0) = 0$, which assumption we make.

To continue with the proof we initially assume that our F is twice continuously differentiable on $\mathbf{R}^+ \times \mathbf{R}^+$. Then, as noted above, assumption (1) is equivalent to the condition $F_{12} \geq 0$, and we can write

$$F(u, v) = \int_0^\infty \int_0^\infty F_{12}(s, t) \Theta(u - s) \Theta(v - t) d\mathcal{L}^1 s d\mathcal{L}^1 t,$$

where $\Theta(r) = 1$ for $r > 0$ and $\Theta(r) = 0$ for $r \leq 0$. Hence,

$$\begin{aligned} & \iint F(f(x), g(y))W(x-y) d\mathcal{L}^n x d\mathcal{L}^n y \\ &= \iint F_{12}(s, t) \left[\iint \chi_{\{f>s\}}(x) \chi_{\{g>t\}}(y) W(x-y) d\mathcal{L}^n x d\mathcal{L}^n y \right] d\mathcal{L}^1 s d\mathcal{L}^1 t. \end{aligned}$$

The standard Riesz rearrangement inequality noted above tells us that the inner double integral does not decrease if we replace $\chi_{\{f>s\}}$ by $(\chi_{\{f>s\}})^* = \chi_{\{f^*>s\}}$, replace $\chi_{\{g>t\}}$ by $(\chi_{\{g>t\}})^* = \chi_{\{g^*>t\}}$, and replace W by W^* . This completes the proof of inequality (2) in case F has two derivatives.

Now assume only that F is merely continuous. For $\varepsilon > 0$ we set

$$f_\varepsilon = \min\{\varepsilon^{-1}, \max\{(f - \varepsilon), 0\}\} \quad \text{and} \quad g_\varepsilon = \min\{\varepsilon^{-1}, \max\{(g - \varepsilon), 0\}\}.$$

If we could prove inequality (2) for all such f_ε 's and g_ε 's, then (2) would follow for our original f and g by Lebesgue's monotone convergence theorem—note that (1) implies that $F(u, v)$ is monotone in u and v . Since f and g are in $\mathcal{E}_0$, the function $\chi_{\{f>\varepsilon\}}(x) \chi_{\{g>\varepsilon\}}(y) W(x-y)$ belongs to $\mathbf{L}^1(\mathbf{R}^n \times \mathbf{R}^n)$. In view of our remarks above, it is sufficient to prove inequality (2) under the assumption that $F(0, v) = F(u, 0) = 0$, which assumption we make. We now extend F to a function $G: \mathbf{R}^2 \rightarrow \mathbf{R}^+$ by setting $G(u, v) = F(u, v)$ in case $u \geq 0$ and $v \geq 0$, and $G(u, v) = 0$ otherwise. It is immediate to check that G also satisfies (1) whenever $u_2 \geq u_1 > -\infty$ and $v_2 \geq v_1 > -\infty$. For small positive numbers δ , we let Φ_δ be a standard nonnegative infinitely differentiable mollifying function on $\mathbf{R}^2$ with support contained in a δ -neighborhood of the origin and set $F_\delta = G * \Phi_\delta - (G * \Phi_\delta)(0, 0)$; here $*$ denotes convolution. It is then immediate to check that $F_\delta \geq 0$ on $\mathbf{R}^+ \times \mathbf{R}^+$ with $F_\delta(0, 0) = 0$ and that F_δ also satisfies (1). Furthermore, each F_δ is infinitely differentiable and, for fixed ε , F_δ will be uniformly very close to F on the domain $[0, \varepsilon^{-1}] \times [0, \varepsilon^{-1}]$ provided δ is very small. Our argument above shows that inequality (2) holds for f_ε , g_ε , and each F_δ . We let $\delta \rightarrow 0$ to infer that inequality (2) holds for f_ε , g_ε , and F . In view of our remark above, the general inequality (2) then follows.

Our final inequality asserted in the theorem is the limiting form of (2) when W is a delta function. One repeats the proof of (2) *mutatis mutandis*. ■

Remarks. (1) The case $F(u, v) = uv$ gives Riesz's original theorem.

(2) Theorem 2.2 is easily extended to apply to $p > 2$ functions and to $k > 2$ variables in $\mathbf{R}^n$ based on the generalization of Riesz's inequality given in [BLL].

2.3. Corollary (A rearrangement inequality for convex integrands). Suppose $\Psi: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is convex with $\Psi(0) = 0$. Suppose also that f, g belong to our class $\mathcal{E}_0$ and $\Psi \circ f$ and $\Psi \circ g$ both belong to $\mathbf{L}^1(\mathbf{R}^n)$. Then

$$\begin{aligned} & \iint \Psi(|f(x) - g(y)|) W(ax + by) d\mathcal{L}^n x d\mathcal{L}^n y \\ & \geq \iint \Psi(|f^*(x) - g^*(y)|) W^*(ax + by) d\mathcal{L}^n x d\mathcal{L}^n y \end{aligned}$$

and

$$\int \Psi(|f(x) - g(x)|) d\mathcal{L}^n x \geq \int \Psi(|f^*(x) - g^*(x)|) d\mathcal{L}^n x$$

for every nonnegative function W in $L^1(\mathbf{R}^n)$ and every choice of nonzero numbers a and b .

Proof. The corollary will be proved if we prove that the function $F(u, v) = \Psi(u) + \Psi(v) - \Psi(|u - v|)$ satisfies condition (1) of Theorem 2.2. Note that $F(u, v) \geq 0$ since Ψ is nondecreasing. Note also that the introduction here of $\Psi(u) + \Psi(v)$ forces us to assume $\Psi \circ f, \Psi \circ g \in L^1(\mathbf{R}^n)$. We define $\tilde{\Psi}(t) = \Psi(|t|)$ for $t \in \mathbf{R}$. Then $\tilde{\Psi}$ is convex and (1) reads $\tilde{\Psi}(\alpha) + \tilde{\Psi}(\beta) \leq \tilde{\Psi}(\gamma) + \tilde{\Psi}(\delta)$ with $\alpha = v_2 - u_2$, $\beta = v_1 - u_1$, $\gamma = v_2 - u_1$, $\delta = v_1 - u_2$, so that $\alpha + \beta = \gamma + \delta$. Since $v_2 \geq v_1$ and $u_2 \geq u_1$, we have that γ is the largest and δ is the smallest of these four numbers. Inequality (1) is then satisfied as a consequence of the convexity of $\tilde{\Psi}$. ■

Remarks. (1) The second inequality in Corollary 2.3 was proved by Chiti [CG] and by Crandall and Tartar [CT]. See [KB, p. 23].

(2) There is an interesting limiting case of Theorem 2.2. By a monotone convergence argument, we can take $W(x) = 1$ for each x . Then

$$\iint F(f(x), g(y)) d\mathcal{L}^n x d\mathcal{L}^n y \leq \iint F(f^*(x), g^*(y)) d\mathcal{L}^n x d\mathcal{L}^n y,$$

(3) The second inequality in Corollary 2.3 asserts that the rearrangement map is a contraction for the map $(f, g) \rightarrow \int \Psi(|f - g|) d\mathcal{L}^n$ and therefore settles the question about the convergence of f_j^* to f^* in the following sense. Suppose f and the f_j 's belong to our class $\mathcal{E}_0$ and that $\Psi: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is convex with $\Psi(0) = 0$. If $\Psi \circ f$ and the $\Psi \circ f_j$'s all belong to $L^1(\mathbf{R}^n)$ and

$$\text{if } \int \Psi(|f_j - f|) d\mathcal{L}^n \rightarrow 0, \text{ then } \int \Psi(|f_j^* - f^*|) d\mathcal{L}^n \rightarrow 0 \text{ as } j \rightarrow \infty.$$

Trivially $\Psi \circ f^*$ and $\Psi \circ f_j^*$ are in $L^1(\mathbf{R}^n)$ because, for example, $\|\Psi \circ f^*\|_{L^1} = \|\Psi \circ f\|_{L^1}$.

Another convergence result we shall need for the proof of our main theorem, Theorem 4.1, concerns convergence in measure. The following theorem can easily be extended to functions from $\mathbf{R}^n$ to $\mathbf{C}$ but, for simplicity, we state and prove it only for functions from $\mathbf{R}^n$ to $\mathbf{R}^+$.

2.4. Theorem (Convergence in measure). *Let $f, f_1, f_2, f_3, \dots$ be functions in our class $\mathcal{E}_0$ such that $f_j \rightarrow f$ in measure as $j \rightarrow \infty$; this means for each $\varepsilon > 0$ there is some $J_\varepsilon < \infty$ such that $\mathcal{L}^n\{x: |f_j(x) - f(x)| > \varepsilon\} < \varepsilon$ whenever $j > J_\varepsilon$. Then also $f_j^* \rightarrow f^*$ in measure as $j \rightarrow \infty$.*

Proof. Step 1. We first prove the theorem assuming the existence of numbers Y and β such that $f_{(j)}(x) \leq Y$ for each x and the sets $K_{(j)} = \{x: f_{(j)}(x) > 0\}$ satisfy the inequality $\mathcal{L}^n(K_{(j)}) \leq \beta$. Here the symbol (j) means either j or

no j . If $j > J_\varepsilon$, then by assumption $|f_j(x) - f(x)| \leq \varepsilon$ except for x 's in some set $B_j(\varepsilon)$ with $\mathcal{L}^n(B_j(\varepsilon)) < \varepsilon$. Hence,

$$\int |f_j - f| d\mathcal{L}^n \leq \varepsilon \mathcal{L}^n(K) + \varepsilon \mathcal{L}^n(K_j) + Y \mathcal{L}^n(B_j(\varepsilon)) < \varepsilon(2\beta + Y).$$

Since ε is arbitrary, we infer that $f_j \rightarrow f$ in $L^1(\mathbf{R}^n)$ norm as $j \rightarrow \infty$. Corollary 2.3 implies that $f_j^* \rightarrow f^*$ in $L^1(\mathbf{R}^n)$ norm as well, and this, as is well known, implies that $f_j^* \rightarrow f^*$ in measure.

Step 2. We turn now to the general case. The failure of our theorem would imply the existence of positive numbers ε_0 and γ such that

$$\mathcal{L}^n\{x : |f_j^*(x) - f^*(x)| > 3\varepsilon_0\} > 4\gamma$$

for some subsequence of the f_j 's. We will show that this is not possible and thereby prove the theorem. We pass to such a subsequence, which will still be denoted by f_j . We now fix a number $Y > \varepsilon_0$ such that

$$\mathcal{L}^n\{x : f(x) > Y - \varepsilon_0\} < \gamma,$$

and, in terms of ε_0 and this Y , define three operators τ^1, τ^2, τ^3 on measurable functions $g : \mathbf{R}^n \rightarrow \mathbf{R}^+$ by setting

$$\begin{aligned} (\tau^1 g)(x) &= \min\{g(x), \varepsilon_0\}, \\ (\tau^2 g)(x) &= \begin{cases} 0 & \text{if } 0 \leq g(x) \leq \varepsilon_0, \\ g(x) - \varepsilon_0 & \text{if } \varepsilon_0 \leq g(x) \leq Y, \\ Y - \varepsilon_0 & \text{if } Y \leq g(x), \end{cases} \\ (\tau^3 g)(x) &= \max\{g(x) - Y, 0\}. \end{aligned}$$

We note the following facts (with $i = 1, 2$, or 3)

- (1) $g = \tau^1 g + \tau^2 g + \tau^3 g$,
- (2) $(\tau^i g)^* = \tau^i(g^*)$,
- (3) $|\tau^i g - \tau^i h| \leq |g - h|$.

(3) can be proved by considering the various cases in the definition of the τ^i 's, e.g., nine cases when $i = 2$.

Our next goal will be to show that $\tau^2 f_j \rightarrow \tau^2 f$ in measure. For each $\varepsilon > 0$, define the sets

$$A_j^i(\varepsilon) = \{x : |\tau^i f_j(x) - \tau^i f(x)| > \varepsilon\} \quad \text{and} \quad A_j(\varepsilon) = \{x : |f_j(x) - f(x)| > \varepsilon\}$$

for each $i = 1, 2, 3$ and each j . (1) and the triangle inequality imply

$$|\tau^2 f_j - \tau^2 f| \leq |f_j - f| + |\tau^1 f_j - \tau^1 f| + |\tau^3 f_j - \tau^3 f|.$$

Hence, if $|\tau^2 f_j(x) - \tau^2 f(x)| > 3\varepsilon$ for some x , then at least one of the quantities $|f_j(x) - f(x)|$, $|\tau^1 f_j(x) - \tau^1 f(x)|$, or $|\tau^3 f_j(x) - \tau^3 f(x)|$ must exceed ε . Thus,

$$(4) \quad \mathcal{L}^n(A_j^2(3\varepsilon)) \leq \mathcal{L}^n(A_j(\varepsilon)) + \mathcal{L}^n(A_j^1(\varepsilon)) + \mathcal{L}^n(A_j^3(\varepsilon)) \leq 3\mathcal{L}^n(A_j(\varepsilon));$$

the last inequality is a consequence of (3). Since $f_j \rightarrow f$ in measure, (4) implies that $\tau^2 f_j \rightarrow \tau^2 f$ in measure as well.

Now we wish to estimate the measure of the sets $K_{(j)} = \{x : \tau^2 f_{(j)}(x) > 0\}$. Let $L = \{x : f(x) > \varepsilon_0/2\} \supset K$. Since $f \in \mathcal{E}_0$, $\mathcal{L}^n(L) < \infty$. If $x \in K_j$ but $x \notin L$, then $|f_j(x) - f(x)| > \varepsilon_0/2$ and thus, in case $j > J_{\varepsilon_0/2}$, we have $\mathcal{L}^n\{x : x \in K_j \text{ and } x \notin L\} < \varepsilon_0/2$. Therefore, for such j 's, $\mathcal{L}^n(K_j) \leq \mathcal{L}^n(L) + \varepsilon_0/2 \equiv \beta$. Also $\mathcal{L}^n(K) \leq \mathcal{L}^n(L) < \beta$. Therefore, the functions $\tau^2 f, \tau^2 f_1, \tau^2 f_2, \tau^2 f_3, \dots$ satisfy the hypotheses of Step 1 and we conclude, using (2), that $\tau^2(f_j^*) \rightarrow \tau^2(f^*)$ in measure.

For each i and j , we set

$$C_j \equiv \{x : |f_j^*(x) - f^*(x)| > 3\varepsilon_0\}, \quad C_j^i = \{x : |\tau^i f_j^*(x) - \tau^i f^*(x)| > \varepsilon_0\}.$$

Our initial assumption in Step 2 about the failure of the theorem is that $\mathcal{L}^n(C_j) > 4\gamma$ for each j in our present subsequence. We infer from the triangle inequality that for each j ,

$$|f_j^* - f^*| \leq |\tau^1 f_j^* - \tau^1 f^*| + |\tau^2 f_j^* - \tau^2 f^*| + |\tau^3 f_j^* - \tau^3 f^*|$$

which in turn implies that $C_j \subset C_j^1 \cup C_j^2 \cup C_j^3$ which then further implies $\mathcal{L}^n(C_j) \leq \mathcal{L}^n(C_j^1) + \mathcal{L}^n(C_j^2) + \mathcal{L}^n(C_j^3)$. Since $\tau^2 f_j^* \rightarrow \tau^2 f^*$ in measure, we know that $\mathcal{L}^n(C_j^2) < \gamma$ for all sufficiently large j . We also know that each C_j^1 is empty because $0 \leq \tau^1(f_{(j)}^*)(x) = (\tau^1 f_{(j)})^*(x) \leq \varepsilon_0$ for each x . To complete the proof, it suffices to show that $\mathcal{L}^n(C_j^3) < 3\gamma$ for large enough j . Clearly,

$$\begin{aligned} \mathcal{L}^n(C_j^3) &\leq \mathcal{L}^n\{x : f^*(x) > Y\} + \mathcal{L}^n\{x : f_j^*(x) > Y\} \\ &= \mathcal{L}^n\{x : f(x) > Y\} + \mathcal{L}^n\{x : f_j(x) > Y\}. \end{aligned}$$

We have $\mathcal{L}^n\{x : f(x) > Y\} \leq \mathcal{L}^n\{x : f(x) > Y - \varepsilon_0\} < \gamma$ from the definition of Y . Since $f_j \rightarrow f$ in measure, we infer the existence of sets D_j with $\lim_{j \rightarrow \infty} \mathcal{L}^n(D_j) = 0$ such that $\{x : f_j(x) > Y\} \subset \{x : f(x) > Y - \varepsilon_0\} \cup D_j$. Thus, $\mathcal{L}^n(C_j^3) \leq \gamma + \gamma + \mathcal{L}^n(D_j)$, and this right side is less than 3γ for all sufficiently large j . ■

Having disposed of the question of the convergence of f_j^* to f^* , we turn now to the much more interesting question, which is the central topic of this paper, the convergence of the gradients. While our primary concern is with functions in class $\mathcal{E}_1$ (because, for example, nonnegative functions in $\mathbf{W}^{1,p}$ belong to $\mathcal{E}_1$ but are not necessarily in $\mathcal{E}_2$ when $p > 1$), we first analyze the smaller class $\mathcal{E}_2$. In §§7 and 8, we shall show how to reduce everything to questions about $\mathcal{E}_2$. The first interesting fact about $\mathcal{E}_2$ is the following.

2.5. Theorem (Functions in $\mathcal{E}_2$ belong to $\mathbf{L}^p$). *Set $p = \infty$ in case $n = 1$ and $p = n/(n - 1)$ otherwise. Suppose $f : \mathbf{R}^n \rightarrow \mathbf{R}$ is $\mathcal{L}^n$ measurable and $|f|$*

belongs to $\mathcal{E}_2$. Then f also belongs to $L^p(\mathbf{R}^n)$ with $\|f\|_{L^p} \leq C\|\nabla f\|_{L^1} = C\|(|\nabla f|)\|_{L^1}$ for a suitable constant C . Furthermore, such an f is also continuous in case $n = 1$. Consequently, if $f, f_1, f_2, f_3, \dots : \mathbf{R}^n \rightarrow \mathbf{R}$ are $\mathcal{L}^n$ measurable and $|f|, |f_1|, |f_2|, |f_3|, \dots$ all belong to $\mathcal{E}_2$ and if the ∇f_j 's converge to ∇f in $L^1(\mathbf{R}^n, \mathbf{R}^n)$, then the f_j 's converge to f in $L^p(\mathbf{R}^n)$.

Proof. We use a method which can be found after Theorem 2.1 in [BL]. Let $g : \mathbf{R}^n \rightarrow [0, 1]$ be a smooth function with compact support which takes the value 1 in a neighborhood of the origin and whose gradient is bounded in norm by 1. For positive integers k , set $g_k(x) = g(x/k)$. For small positive numbers ε and positive integers k , set

$$f_\varepsilon(x) = \chi_{\{|f|>\varepsilon\}}(|f(x)| - \varepsilon) = (|f(x)| - \varepsilon)_+ \quad \text{and} \quad h_{\varepsilon,k} = g_k f_\varepsilon.$$

It is well known that

$$\nabla h_{\varepsilon,k} = g_k \chi_{\{|f|>\varepsilon\}} \nabla |f| + f_\varepsilon \nabla g_k \quad \text{and} \quad |\nabla f| = \pm \nabla f$$

almost everywhere in the distribution sense. The fact that $|\nabla f| = \pm \nabla f$ implies that $\nabla f \in L^1(\mathbf{R}^n, \mathbf{R}^n)$. Now suppose that $f \in L^\infty(\mathbf{R}^n)$, i.e., there is a constant A with $|f(x)| \leq A$ for almost every x . Since each $h_{\varepsilon,k}$ belongs to $W^{1,1}(\mathbf{R}^n)$, the usual Sobolev inequality implies that

$$\begin{aligned} \|h_{\varepsilon,k}\|_{L^p} &\leq C\|\nabla h_{\varepsilon,k}\|_{L^1} \\ &\leq C\{\| |\nabla f| \|_{L^1} + \|f_\varepsilon \nabla g_k\|_{L^1}\} \\ &\leq C\{\| |\nabla f| \|_{L^1} + AD_\varepsilon/k\}, \end{aligned}$$

where C is another constant and $D_\varepsilon = \|\chi_{\{|f|>\varepsilon\}}\|_{L^1} < \infty$. Letting $k \rightarrow \infty$, we have by Fatou's lemma that $\|f_\varepsilon\|_{L^p} \leq C\|\nabla f\|_{L^1}$. Letting $\varepsilon \rightarrow 0$ and using Fatou's lemma again, we conclude that $\|f\|_{L^p} \leq C\|\nabla f\|_{L^1}$. If $f \notin L^\infty(\mathbf{R}^n)$, choose some large number T and set $f_T(x) = \min\{f(x), T\}$. Then $\nabla f_T = \chi_{\{f < T\}} \nabla f$. Since $f_T \in L^\infty(\mathbf{R}^n)$, the preceding estimate implies that f_T is uniformly (in T) bounded in L^p . The desired conclusion follows by Fatou's lemma. The fact that f is continuous in case $n = 1$ is a standard fact about Sobolev imbeddings. ■

The next two observations concern rearrangement of functions in $\mathcal{E}_2$.

2.6. Theorem (Rearrangement preserves $\mathcal{E}_2$). *If $f \in \mathcal{E}_2$, then $f^* \in \mathcal{E}_2$ with*

$$\int |\nabla f^*| d\mathcal{L}^n \leq \int |\nabla f| d\mathcal{L}^n.$$

2.7. Theorem (Rearrangement decreases integrals of convex integrands). *Suppose f belongs to $\mathcal{E}_1$. Then $f^* \in \mathcal{E}_1$ (by applying Theorem 2.6 to the functions $\max\{f - \varepsilon, 0\}$ for each ε). If Ψ is a suitable integrand and if $\Psi \circ |\nabla f|$ belongs to $L^1(\mathbf{R}^n)$, then $\Psi \circ |\nabla f^*|$ also belongs to $L^1(\mathbf{R}^n)$ with*

$$\int \Psi(|\nabla f^*|) d\mathcal{L}^n \leq \int \Psi(|\nabla f|) d\mathcal{L}^n.$$

The primary assertion in Theorem 2.6 is that ∇f^* is a function. (The secondary fact that ∇f^* is in $L^1(\mathbf{R}^n)$ is also covered by Theorem 2.7.) Since f^* is a radial function, its gradient is purely radial, but this gradient could, in principal, be a Radon measure that is not absolutely continuous with respect to $\mathcal{L}^1$. Theorem 2.6 was first proved for f 's in $\mathbf{W}^{1,1}$ by Hilden [HK]. It was also proved by Brothers and Ziemer [BZ] when f has compact support. To prove it for f in $\mathcal{E}_2$, set, for small positive numbers ε , $f_\varepsilon = (f - \varepsilon)_+$. Clearly, $f_\varepsilon \in \mathbf{W}^{1,1}$ (because $f \in L^p(\mathbf{R}^n)$ with $p = n/(n-1)$) and $f_\varepsilon^* = (f^* - \varepsilon)_+$. Moreover, $\nabla(f - \varepsilon)_+ = \chi_{\{f > \varepsilon\}} \nabla f$. By Hilden's theorem, f_ε^* belongs to $\mathbf{W}^{1,1}$ and

$$\begin{aligned} \int \chi_{\{f > \varepsilon\}} |\nabla f| d\mathcal{L}^n &= \int |\nabla f_\varepsilon| d\mathcal{L}^n \\ &\geq \int |\nabla f_\varepsilon^*| d\mathcal{L}^n = \int \chi_{\{f > \varepsilon\}} |\nabla f^*| d\mathcal{L}^n. \end{aligned}$$

Theorem 2.6 then follows by monotone convergence.

Theorem 2.7 was proved for $\mathbf{W}^{1,p}(\mathbf{R}^n)$ by Sperner [SE1, SE2] and by Talenti [TG]; see also [LE1] for the $\mathbf{W}^{1,2}$ case. The case of general convex Ψ was proved for analytic f by Bandle [BC] and, more generally, by Brothers and Ziemer [BZ] for f having compact support. We were informed by Brothers that their proof extends to functions in our class $\mathcal{E}_1$.

2.8. Extension of symmetric decreasing rearrangement to vector-valued functions. In Definition 1.3, the symmetric decreasing rearrangement operator $\mathcal{R}$ was defined for functions from $\mathbf{R}^n \rightarrow \mathbf{R}^+ = \{x : 0 \leq x < \infty\}$. Here we show how to extend the definition to functions from $\mathbf{R}^n$ to $\mathbf{R}$ or from $\mathbf{R}^n$ to $\mathbf{C}$ or, most generally, from $\mathbf{R}^n$ to $\mathbf{R}^m$. The extension is easy. If $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$, then by definition

$$f^* = \mathcal{R}f \equiv |f|^*.$$

More precisely,

$$|f|(x) = \left(\sum_{\alpha=1}^m (f^\alpha(x))^2 \right)^{1/2}$$

for each x . If $f : \mathbf{R}^n \rightarrow \mathbf{C}$, then $\mathbf{C}$ is regarded as $\mathbf{R}^2$, and $|f|$ agrees with the usual complex norm.

We say that a function $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$ belongs to (our new) $\mathcal{E}_1$ (resp. belongs to (our new) $\mathcal{E}_2$) if and only if f is $\mathcal{L}^n$ measurable and the norm of each component belongs to $\mathcal{E}_1$ (resp. belongs to $\mathcal{E}_2$), i.e., $|f^i| \in \mathcal{E}_1$ (resp. $|f^i| \in \mathcal{E}_2$) ($i = 1, \dots, m$). This implies, in particular, that ∇f^i belongs to $L^1(\Omega)$ for each measurable subset Ω of $\mathbf{R}^n$ having finite measure ($i = 1, \dots, m$).

Suppose now that $f, f_1, f_2, \dots : \mathbf{R}^n \rightarrow \mathbf{R}^m$ is a sequence of functions belonging to $\mathcal{E}_1$ with $f_j \rightarrow f$ and $\nabla f_j \rightarrow \nabla f$ in some suitable sense as $j \rightarrow \infty$. If we can use this assumed convergence to show that $|f_j| \rightarrow |f|$ and $\nabla |f_j| \rightarrow \nabla |f|$ in a suitable sense, then the theorems in the rest of this paper can be used to

decide whether or not f_j converges to f^* —such convergence is assured if $|f|$ is co-area regular.

The convergence of the $|f_j|$'s to $|f|$ can be assured, as we now prove, if the f_j 's converge to f in the following sense. We assume the following two conditions.

(i) For each subsequence of the f_j 's, there exists a further subsubsequence in which the functions f_j converge to f pointwise $\mathcal{L}^n$ almost everywhere. This condition is natural from the point of view of, say, $\mathbf{W}^{1,p}(\mathbf{R}^n)$ for which such required convergence is well known.

(ii) There is a suitable integrand Ψ such that $\Psi \circ |\nabla f|$ belongs to $\mathbf{L}^1(\mathbf{R}^n)$ and

$$(1) \quad \int \Psi(|\nabla f_j - \nabla f|) d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

Here for $u : \mathbf{R}^n \rightarrow \mathbf{R}^m$, we write

$$|\nabla u(x)|^2 = \sum_{i=1}^n \sum_{\alpha=1}^m \left(\frac{\partial u^\alpha}{\partial x^i}(x) \right)^2.$$

Assuming that (i) and (ii) hold (and that all our functions are in $\mathcal{E}_1$), we will prove that (i) and (ii) hold for the functions $|f|, |f_1|, |f_2|, |f_3|, \dots$. Of course, (i) is trivial so we concentrate on (ii). It is well known that for each $i = 1, \dots, m$ and for almost every x ,

$$\frac{\partial |f|}{\partial x^i}(x) = \begin{cases} \frac{1}{|f(x)|} \sum_{\alpha=1}^m f^\alpha(x) \frac{\partial f^\alpha}{\partial x^i}(x) & \text{if } |f(x)| > 0, \\ 0 & \text{otherwise.} \end{cases}$$

We are, of course, using the fact that each ∇f^α vanishes almost everywhere on the set on which f vanishes (see Theorem 3.1). This formula implies the existence of a function $\theta : \mathbf{R}^n \rightarrow \mathbf{S}^{m-1}$ such that for each i and almost every x ,

$$(2) \quad \frac{\partial |f|}{\partial x^i}(x) = \sum_{\alpha=1}^m \theta^\alpha(x) \frac{\partial f^\alpha}{\partial x^i}(x).$$

Using (2) for $|f|$ and the $|f_j|$'s, we obtain for each j ,

$$\begin{aligned} |\nabla |f_j| - \nabla |f||^2 &= \sum_{i=1}^n \left(\sum_{\alpha=1}^m \theta_j^\alpha \frac{\partial f_j^\alpha}{\partial x^i} - \theta^\alpha \frac{\partial f^\alpha}{\partial x^i} \right)^2 \\ &\leq m \sum_{i=1}^n \sum_{\alpha=1}^m \left(\theta_j^\alpha \frac{\partial f_j^\alpha}{\partial x^i} - \theta^\alpha \frac{\partial f^\alpha}{\partial x^i} \right)^2 \\ &\leq 2m \sum_{i=1}^n \sum_{\alpha=1}^m \left(\left[(\theta_j^\alpha - \theta^\alpha) \frac{\partial f^\alpha}{\partial x^i} \right]^2 + \left[\theta_j^\alpha \frac{\partial (f_j^\alpha - f^\alpha)}{\partial x^i} \right]^2 \right). \end{aligned}$$

Now we use the facts that $|\theta_j^\alpha| \leq 1$ and that Ψ is increasing together with the condition that $\Psi(2t) \leq C\Psi(t)$ to conclude the existence of a constant D such that

$$(3) \quad \Psi(|\nabla|f_j| - \nabla|f||) \leq D \sum_{i=1}^n \sum_{\alpha=1}^m \left(\Psi \left(\left| (\theta_j^\alpha - \theta^\alpha) \frac{\partial f^\alpha}{\partial x^i} \right| \right) + \Psi \left(\left| \frac{\partial(f_j^\alpha - f^\alpha)}{\partial x^i} \right| \right) \right).$$

As a consequence of (1) and the monotonicity of Ψ , we infer that

$$\int \Psi \left(\left| \frac{\partial(f_j^\alpha - f^\alpha)}{\partial x^i} \right| \right) d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty$$

which takes care of the right-hand summand of (3). It remains to be shown for each i and α that

$$(4) \quad \int \Psi \left(\left| (\theta_j^\alpha - \theta^\alpha) \frac{\partial f^\alpha}{\partial x^i} \right| \right) d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

Note that the integrand in (4) belongs to $\mathbf{L}^1(\mathbf{R}^n)$ because $|(\theta_j^\alpha - \theta^\alpha)| \leq 2$ and $\Psi \circ |\partial f^\alpha / \partial x^i|$ belongs to $\mathbf{L}^1(\mathbf{R}^n)$. The integrand is therefore bounded by a constant times $\Psi(|\partial f^\alpha / \partial x^i|)$. To conclude our proof as a consequence of dominated convergence and the continuity of $t \rightarrow \Psi(|t|)$, it is sufficient to establish that

$$(5) \quad (\theta_j^\alpha - \theta^\alpha) \frac{\partial f^\alpha}{\partial x^i} \rightarrow 0 \quad \text{almost everywhere.}$$

To prove (5), we distinguish two sets:

$$A = \{x : f(x) = 0\} \quad \text{and} \quad B = \{x : f(x) \neq 0\}.$$

On A , each $\partial f^\alpha / \partial x^i$ vanishes almost everywhere so that its contribution to (5) is trivial. On B , $f_j^\alpha \rightarrow f^\alpha \neq 0$ and hence $\theta_j^\alpha(x) \rightarrow \theta^\alpha(x)$ since $\theta_j^\alpha(x) = f_j^\alpha(x)/|f_j(x)|$ for each j and $\theta^\alpha(x) = f^\alpha(x)/|f^\alpha(x)|$; here we are invoking condition (i) above.

Our conclusion then is that *with our definition $f^* \equiv |f|^*$, the main theorems in this paper are valid.* We remark that condition (i) above can be replaced by other conditions such as

$$(i)' \quad \int \tilde{\Psi}(|f_j - f|) d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty \quad \text{for some suitable integrand } \tilde{\Psi}.$$

§3. PROPERTIES OF THE CO-AREA DISTRIBUTION FUNCTION $\mathcal{F}_f$

Our goal here is to show that co-area distribution functions $\mathcal{F}_f$ (in Definition 1.2.5) are absolutely continuous for each f and their derivatives $\mathcal{F}_f'$ behave upper semicontinuously with respect to convergence of f 's; this is Theorem 3.5. We begin with Theorem 3.1 which is a generalization of the well-known textbook theorem that ∇f vanishes almost everywhere on each level set of f . Theorem 3.1 asserts the same conclusion on the inverse image of any Borel set of zero $\mathcal{L}^1$ measure—this theorem should be part of the basic literature, but we could not find it there.

3.1. Theorem (Gradients vanish on the inverse image of small sets). *Suppose $f: \mathbf{R}^n \rightarrow \mathbf{R}$ belongs to $L^1_{\text{loc}}(\mathbf{R}^n)$ with distributional gradient ∇f which is a function in $L^1_{\text{loc}}(\mathbf{R}^n, \mathbf{R}^n)$. Suppose also that A is a Borel subset of the line $\mathbf{R}$ having zero $\mathcal{L}^1$ measure, and set $B = f^{-1}(A) = \{x: f(x) \in A\} \subset \mathbf{R}^n$. Then B is $\mathcal{L}^n$ measurable (since A is a Borel set) and ∇f vanishes almost everywhere on B .*

Proof. Let $U_1 \supset U_2 \supset U_3 \supset \cdots$ be open subsets of the line, each containing A , such that $\mathcal{L}^1(U_j) \rightarrow 0$ as $j \rightarrow \infty$, and set $V = \bigcap_{j=1}^{\infty} U_j \supset A$. Now define

$$M_j(t) = \int_0^t \chi_{U_j}(s) d\mathcal{L}^1 s \quad \text{for each } t \in \mathbf{R},$$

noting that $M_j(t) \leq |t|$ for each t and that $M_j(t) \rightarrow 0$ uniformly in t as $j \rightarrow \infty$. We assert that

$$(1) \quad \int [f - M_j \circ f] \operatorname{div} \phi d\mathcal{L}^n = - \int [1 - \chi_{\{f \in U_j\}}](\nabla f, \phi) d\mathcal{L}^n$$

for each j and for every function $\phi \in \mathcal{C}_c^\infty(\mathbf{R}^n, \mathbf{R}^n)$.

We will first show that (1) implies the assertion of our theorem. Since $|M_j \circ f| \leq |f|$ and $(M_j \circ f)(x) \rightarrow 0$ as $j \rightarrow \infty$ for every $x \in \mathbf{R}^n$, Lebesgue's dominated convergence theorem implies that the left side of (1) converges to $D_\phi \equiv \int f \operatorname{div} \phi d\mathcal{L}^n$ as $j \rightarrow \infty$. Since $0 \leq 1 - \chi_{\{f \in U_j\}} \leq 1$ and since for each $x \in \mathbf{R}^n$, $\chi_{\{f \in U_j\}}(x) \rightarrow \chi_{\{f \in V\}}(x)$, the right side of (1) converges (by dominated convergence) to $E_\phi = - \int [1 - \chi_{\{f \in V\}}](\nabla f, \phi) d\mathcal{L}^n$ as $j \rightarrow \infty$. Therefore, $D_\phi = E_\phi$ which means that

$$- \int (\nabla f, \phi) d\mathcal{L}^n = \int f \operatorname{div} \phi d\mathcal{L}^n = - \int [1 - \chi_{\{f \in V\}}](\nabla f, \phi) d\mathcal{L}^n.$$

This holds for every ϕ , and hence the uniqueness of the distributional derivative ∇f implies that

$$\nabla f = [1 - \chi_{\{f \in V\}}] \nabla f$$

$\mathcal{L}^n$ almost everywhere. Thus, $\nabla f(x) = 0$ for almost every $x \in f^{-1}(V)$, which proves the theorem since $A \subset V$.

Now we turn to the proof of (1). We will show that for any open subset U of the line $\mathbf{R}$ having finite one-dimensional measure,

$$(2) \quad \int [f - M \circ f] \operatorname{div} \phi d\mathcal{L}^n = \int [1 - \chi_{\{f \in U\}}](\nabla f, \phi) d\mathcal{L}^n;$$

here

$$M(t) = \int_0^t \chi_U(s) d\mathcal{L}^1 s \quad \text{for each } t \in \mathbf{R}.$$

We let $0 \leq g_1 \leq g_2 \leq g_3 \leq \cdots$ be continuous functions on the line $\mathbf{R}$ which converge pointwise to χ_U , and for each positive integer j , set

$$N_j(t) = \int_0^t g_j(s) d\mathcal{L}^1 s \quad \text{for each } t \in \mathbf{R}.$$

Since each N_j is differentiable with $0 \leq N'_j = g_j \leq 1$, we know in the sense of distributions that

$$\nabla(N_j \circ f) = (N'_j \circ f)\nabla f = (g_j \circ f)\nabla f.$$

Thus,

$$\int [f - N_j \circ f] \operatorname{div} \phi \, d\mathcal{L}^n = - \int [1 - g_j \circ f](\nabla f, \phi) \, d\mathcal{L}^n.$$

As $j \rightarrow \infty$, $N_j(t) \rightarrow M(t)$ for each t (by dominated convergence since $g_j \rightarrow \chi_U$), and hence

$$\int [f - N_j \circ f] \operatorname{div} \phi \, d\mathcal{L}^n \rightarrow \int [f - M \circ f] \operatorname{div} \phi \, d\mathcal{L}^n$$

by dominated convergence (since $|N_j \circ f| \leq |f|$). Likewise, $g_j \circ f \rightarrow \chi_{\{f \in U\}}$ pointwise and

$$\int [1 - g_j \circ f](\nabla f, \phi) \, d\mathcal{L}^n \rightarrow \int [1 - \chi_{\{f \in U\}}](\nabla f, \phi) \, d\mathcal{L}^n$$

by dominated convergence since $0 \leq g_j \circ f \leq 1$. This proves (2). ■

3.2. Theorem ($\mathcal{F}_f$ is absolutely continuous). Let $f: \mathbf{R}^n \rightarrow \mathbf{R}^+$ be a function in our class $\mathcal{C}_1$ and let $\mathcal{F}_f: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ be the co-area distribution function in Definition 1.2.5. Then $\mathcal{F}_f$ is absolutely continuous on (ε, ∞) for each $\varepsilon > 0$. Equivalently, the distribution derivative $\mathcal{F}'_f$ is a (negative) Radon measure on $(0, \infty)$ which is absolutely continuous with respect to Lebesgue measure. Furthermore,

$$\mathcal{F}'_f(A) = - \int \chi_{\{f \in A\}} \chi_{\{\nabla f \neq 0\}} \, d\mathcal{L}^n$$

for every Borel subset A of $\mathbf{R}^+$.

Proof. It follows readily from the definitions of $\mathcal{F}_f$ and the derivative (regarded as a measure) that

$$(1) \quad \mathcal{F}'_f(U) = - \int \chi_{\{f \in U\}} \chi_{\{\nabla f \neq 0\}} \, d\mathcal{L}^n$$

for each bounded open interval U in (ε, ∞) . Lebesgue's monotone convergence theorem then implies that (1) holds for any open subset U of (ε, ∞) (since these are necessarily countable unions of bounded open intervals). If $\varepsilon > 0$ and $U_1 \supset U_2 \supset U_3 \supset \cdots$ are nested bounded open subsets of (ε, ∞) with $V = \bigcap_{i=1}^{\infty} U_i$, then (1) also holds for such V since Lebesgue's dominated convergence theorem implies that

$$\begin{aligned} \mathcal{F}'_f(V) &= \lim_{i \rightarrow \infty} \mathcal{F}'_f(U_i) \\ &= - \lim_{i \rightarrow \infty} \int \chi_{\{f \in U_i\}} \chi_{\{\nabla f \neq 0\}} \, d\mathcal{L}^n \\ &= - \int \chi_{\{f \in V\}} \chi_{\{\nabla f \neq 0\}} \, d\mathcal{L}^n. \end{aligned}$$

Since $\mathcal{F}'_f$ is a regular Borel measure, we know that for each subset A of $(2\varepsilon, \infty)$, there will exist such a $V = \bigcap_{i=1}^{\infty} U_i$ containing A such that

$$\mathcal{F}'_f(A) = \mathcal{F}'_f(V) = - \int \chi_{\{f \in V\}} \chi_{\{\nabla f \neq 0\}} d\mathcal{L}^n.$$

In case such A has zero $\mathcal{L}^1$ measure, then the corresponding V will have zero $\mathcal{L}^1$ measure and Theorem 3.1 then implies that $\mathcal{F}'_f(A) = 0$. We readily infer from this that $\mathcal{F}_f$ is absolutely continuous on (ε, ∞) for each $\varepsilon > 0$. For a general such A which is a Borel set and associated $V = \bigcap_{i=1}^{\infty} U_i$ containing A with $\mathcal{F}'_f(A) = \mathcal{F}'_f(V)$, we know that $V \sim A$ is a Borel set with zero $\mathcal{L}^1$ measure. Hence, we can use Theorem 3.1 again to conclude

$$\begin{aligned} \mathcal{F}'_f(A) &= - \int \chi_{\{f \in V\}} \chi_{\{\nabla f \neq 0\}} d\mathcal{L}^n \\ &= - \int \chi_{\{f \in A\}} \chi_{\{\nabla f \neq 0\}} d\mathcal{L}^n - \int \chi_{\{f \in V \sim A\}} \chi_{\{\nabla f \neq 0\}} d\mathcal{L}^n \\ &= - \int \chi_{\{f \in A\}} \chi_{\{\nabla f \neq 0\}} d\mathcal{L}^n \end{aligned}$$

which is the final assertion of our theorem. ■

The next theorem is needed not only here but in later parts of this paper as well. It is stated with more generality than we actually require in our applications.

3.3. Theorem (Equi-absolute continuity). *Let $f, f_1, f_2, f_3, \dots$ be $\mathcal{L}^n$ measurable functions mapping $\mathbf{R}^n$ to $\mathbf{R}$ such that $|f|, |f_1|, |f_2|, |f_3|, \dots$ all belong to our class $\mathcal{E}_2$. Suppose also that the ∇f_j 's converge strongly to ∇f in $\mathbf{L}^1(\mathbf{R}^n, \mathbf{R}^n)$ norm. According to Theorem 2.5, the f_j 's converge strongly to f in $\mathbf{L}^p(\mathbf{R}^n)$ norm. Now let $\Phi: \mathbf{R}^n \rightarrow \mathbf{R}$ be any Borel function which satisfies the following two conditions for some finite constant C :*

- (i) $|\xi| + |\Phi(\xi)| \leq C(|\eta| + |\Phi(\eta)|)$ whenever $\xi, \eta \in \mathbf{R}^n$ with $|\xi| \leq 2|\eta|$;
- (ii) $|\Phi(\xi)| \leq C|\xi|$ whenever $\xi \in \mathbf{R}^n$ with $|\xi| \leq 1$.

(We do not assume any special monotonicity or continuity properties for Φ .) Assume that the $\Phi \circ \nabla f_j$'s together with $\Phi \circ \nabla f$ all belong to $\mathbf{L}^1(\mathbf{R}^n)$, and assume also that $\Phi \circ (\nabla f_j - \nabla f) \rightarrow 0$ strongly in $\mathbf{L}^1(\mathbf{R}^n)$ as $j \rightarrow \infty$.

Using these functions, we can define a new function $H_j: \mathbf{R}^+ \rightarrow \mathbf{R}$ for each j by setting

$$H_j(y) = \int \chi_{\{|f_j| > y\}} \Phi(\nabla f_j) d\mathcal{L}^n.$$

Our conclusions are as follows.

(1) *For each $\sigma > 0$, the H_j 's are an equi-absolutely continuous family of functions on (σ, ∞) .*

(2) *The numbers $H_j(y)$ converge to zero as $y \rightarrow \infty$, uniformly in j .*

(3) If Φ is continuous, then for each $y \geq 0$,

$$H_j(y) \rightarrow H(y) = \int \chi_{\{|f|>y\}} \Phi(\nabla f) d\mathcal{L}^n \quad \text{as } j \rightarrow \infty.$$

By conclusion (1), this convergence is uniform on compact subsets of $(0, \infty)$.

Proof. Our first conclusion amounts to the following assertion. Suppose $\sigma > 0$ is fixed. Then given any $\varepsilon > 0$, there will be $\delta > 0$ such that whenever N is a positive integer and

$$2\sigma \leq a_1 < b_1 < a_2 < b_2 < \cdots < a_N < b_N < \infty$$

are numbers with

$$\sum_{i=1}^N (b_i - a_i) < \delta,$$

then

$$\sum_{i=1}^N |H_j(b_i) - H_j(a_i)| < \varepsilon \quad \text{for every } j.$$

We will first prove the entire theorem under the additional assumption that (for some possibly new constant C) $|\Phi(\xi)| \leq C|\xi|$ for each $\xi \in \mathbf{R}^n$ (instead of just for ξ 's with $|\xi| \leq 1$). For a_i 's and b_i 's as above, it is easy to construct a continuous function $L: \mathbf{R} \rightarrow [0, 1]$ with compact support in $\{t: |t| > \sigma\}$ for which $L(t) = 1$ whenever $|t|$ belongs to the set $E = \bigcup_{i=1}^N [a_i, b_i]$ and $\int L d\mathcal{L}^1 \leq 4\mathcal{L}^1(E)$. In particular, our sums $\sum_{i=1}^N |H_j(b_i) - H_j(a_i)|$ are not more than $\int |\Phi(\nabla f_j)| L(f_j) d\mathcal{L}^n$. Corresponding to a general continuous function $L: \mathbf{R} \rightarrow [0, 1]$ having compact support in $\{t: |t| > \sigma\}$, we set for each positive integer j ,

$$h_{j,L} = \int |\Phi(\nabla f_j)| L(f_j) d\mathcal{L}^n.$$

In order to prove conclusion (1), it is then sufficient to show that for each $\varepsilon > 0$, there is some $\delta > 0$ such that the number $h_{j,L}$ does not exceed ε for each j provided $\int L d\mathcal{L}^1$ does not exceed δ . For this purpose, we might as well assume that $\Phi(\xi) = C|\xi|$ and then take $C = 1$ which we now do. The negation of the first conclusion of our theorem then implies the existence of some positive number α together with a subsequence of the f_j 's (which we will still refer to as the f_j 's) and a collection of continuous functions $L_j: \mathbf{R} \rightarrow [0, 1]$ each with compact support in $\{t: |t| > \sigma\}$ and for which

$$\int L_j d\mathcal{L}^1 \rightarrow 0 \quad \text{as } j \rightarrow \infty,$$

and

$$(*) \quad h_j \equiv h_{j,L_j} > 4\alpha \quad \text{for each } j.$$

We shall show that this requirement that all h_j 's exceed 4α leads to a contradiction and thus prove conclusion (1).

We now define functions $M_j : \mathbf{R} \rightarrow \mathbf{R}^+$ as indefinite integrals of the L_j 's, i.e.,

$$M_j(y) = \int_{-\infty}^y L_j(t) d\mathcal{L}^1 t \quad \text{for each } y \text{ and each } j.$$

Similarly, we set

$$\varphi_j \equiv M_j \circ f_j : \mathbf{R}^n \rightarrow \mathbf{R}^+.$$

Clearly, $0 \leq \varphi_j \leq \int L_j d\mathcal{L}^1$ so that $\varphi_j \in \mathbf{L}_{\text{loc}}^1(\mathbf{R}^n)$. As is well known, as a distribution $\nabla \varphi_j$ is the $\mathbf{R}^n$ -valued function given by setting for each x ,

$$(\nabla \varphi_j)(x) = M_j'(f_j(x)) \nabla f_j(x) = L_j(f_j(x)) \nabla f_j(x);$$

here M_j' denotes the derivative of M_j . Clearly, $\nabla \varphi_j \in \mathbf{L}^1(\mathbf{R}^n, \mathbf{R}^n)$ with $\|\nabla \varphi_j\|_{\mathbf{L}^1} \leq \|\nabla f_j\|_{\mathbf{L}^1}$.

It is convenient to define functions V and V_j mapping $\mathbf{R}^n$ to $\mathbf{R}^n$ by setting

$$\begin{aligned} V_j(x) &= \frac{\nabla f_j(x)}{|\nabla f_j(x)|} \quad \text{if } \nabla f_j(x) \neq 0, \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

$$\begin{aligned} V(x) &= \frac{\nabla f(x)}{|\nabla f(x)|} \quad \text{if } \nabla f(x) \neq 0, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Then, with $A \cdot B$ denoting the inner product (A, B) ,

$$\int \nabla \varphi_j \bullet V_j d\mathcal{L}^n = h_j > 4\alpha.$$

Since $\nabla f_j \rightarrow \nabla f$ in $\mathbf{L}^1(\mathbf{R}^n, \mathbf{R}^n)$, we can pass to another subsequence (again denoted by f_j 's) such that

$$\begin{aligned} \nabla f_j &\rightarrow \nabla f \quad \text{pointwise almost everywhere,} \\ |\nabla f_j| &\leq G \quad \text{and} \quad |\nabla f| \leq G \end{aligned}$$

for some function $G \in \mathbf{L}^1(\mathbf{R}^n)$. If A denotes the set on which ∇f does not equal zero, we obviously have that

$$V_j \rightarrow V \quad \text{pointwise almost everywhere on } A.$$

We can write

$$h_j = \int_{\mathbf{R}^n} \nabla \varphi_j \bullet V d\mathcal{L}^n + \int_A \nabla \varphi_j \bullet (V_j - V) d\mathcal{L}^n + \int_{\mathbf{R}^n \setminus A} \nabla \varphi_j \bullet (V_j - V) d\mathcal{L}^n.$$

These three integrals are well defined since $|\nabla \varphi_j| \leq |\nabla f_j| \leq G$ and $|V_j| \leq 1$, $|V| \leq 1$. Since $V_j - V \rightarrow 0$ on A and $|\nabla \varphi_j \bullet (V_j - V)| \leq 2G$, we have by Lebesgue's dominated convergence theorem that the second integral goes to zero as $j \rightarrow \infty$. The third integral also goes to zero since $|\nabla \varphi_j| \leq |\nabla f_j| \leq G$,

$\nabla f_j \rightarrow \nabla f$ strongly in $L^1(\mathbf{R}^n)$, ∇f is identically zero on the complement of A , and $|V_j - V| \leq 2$ everywhere. Thus, for large enough j , the number

$$\tilde{h}_j \equiv \int \nabla \varphi_j \bullet V d\mathcal{L}^n$$

is at least equal to 3α .

Since $|\nabla \varphi_j| \leq G$ and $G \in L^1(\mathbf{R}^n)$, there is a large ball in $\mathbf{R}^n$ having characteristic function B such that for all j ,

$$\int |\nabla \varphi_j|(1 - B) d\mathcal{L}^n < \alpha.$$

Thus the numbers

$$\hat{h}_j \equiv \int \nabla \varphi_j \bullet VB d\mathcal{L}^n$$

are at least equal to 2α for large enough j 's. The fixed function $VB : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is in $L^1(\mathbf{R}^n, \mathbf{R}^n) \cap L^\infty(\mathbf{R}^n, \mathbf{R}^n)$. We let K be a fixed, nonnegative smooth function with compact support on $\mathbf{R}^n$ having integral equal to one, and for positive numbers $\varepsilon > 0$ set $K_\varepsilon(x) = \varepsilon^{-n}K(x/\varepsilon)$. It is well known that the convolutions $K_\varepsilon * (VB)$ converge pointwise almost everywhere to VB as $\varepsilon \rightarrow 0$. Since VB has compact support and nowhere exceeds one, each $K_\varepsilon * (VB)$ is a smooth vector field with compact support and with values nowhere exceeding one in norm. Since $|\nabla \varphi_j| < G$, we can estimate

$$\left| \int \nabla \varphi_j \bullet VB d\mathcal{L}^n - \int \nabla \varphi_j \bullet K_\varepsilon * (VB) d\mathcal{L}^n \right| \leq \int G|VB - K_\varepsilon * (VB)| d\mathcal{L}^n.$$

We can thus use Lebesgue's dominated convergence theorem again to conclude the existence of a fixed $\varepsilon_0 > 0$ such that the numbers

$$\hat{h}_{j, \varepsilon_0} \equiv \int \nabla \varphi_j \bullet K_{\varepsilon_0} * (VB) d\mathcal{L}^n$$

are greater than α for all sufficiently large j . Since $K_{\varepsilon_0} * (VB)$ is smooth with compact support, we can integrate by parts to obtain

$$\begin{aligned} \hat{h}_{j, \varepsilon_0} &= - \int \varphi_j \operatorname{div}(K_{\varepsilon_0} * (VB)) d\mathcal{L}^n \\ &\leq M_j(\infty) \int |\operatorname{div}(K_{\varepsilon_0} * (VB))| d\mathcal{L}^n. \end{aligned}$$

But the numbers

$$M_j(\infty) = \int L_j d\mathcal{L}^1$$

converge to zero as $j \rightarrow \infty$ by assumption, thereby establishing our contradiction and hence establishing our first conclusion (under our added assumption).

For the second conclusion, we note that Theorem 2.4 guarantees that $f_j \in L^p(\mathbf{R}^n)$ for each j ; here $p = \infty$ in case $n = 1$, and $p = n/(n - 1)$ otherwise. In fact, $f_j \rightarrow f$ in $L^p(\mathbf{R}^n)$ as $j \rightarrow \infty$. Using Lebesgue's dominated convergence theorem and the fact that $\chi_{\{|f|>y\}}$ converges to zero pointwise almost everywhere as $y \rightarrow \infty$ for any f in $L^p(\mathbf{R}^n)$, we see that each $H_j(y)$

also converges to zero as $y \rightarrow \infty$. The question is the uniformity in j of such convergence. Proceeding as before, we conclude that failure of the desired uniformity would imply the existence of a positive number β together with a sequence of numbers Y_j converging to ∞ as $j \rightarrow \infty$ and associated continuous functions $L_j: \mathbf{R} \rightarrow [0, 1]$ with $L_j(y) = 0$ for $|y| < Y_j$ such that $h_j \equiv h_{j, L_j} > \beta$, at least for a subsequence of the j 's. Passing to a further subsequence if necessary, we can additionally assume that the f_j 's converge to f almost everywhere, in which case the functions $L_j(f_j)$ also converge to zero almost everywhere. Then

$$\begin{aligned} h_j &\leq \int |\nabla f_j| L_j(f_j) d\mathcal{L}^n \\ &\leq \int |\nabla f_j - \nabla f| L_j(f_j) d\mathcal{L}^n + \int |\nabla f| L_j(f_j) d\mathcal{L}^n. \end{aligned}$$

The last integral converges to zero as $j \rightarrow \infty$ by Lebesgue's dominated convergence theorem while the first integral on the right is bounded by $\|\nabla f_j - \nabla f\|_{L^1}$ which also converges to zero. This establishes the second conclusion.

For the third conclusion, we fix $y \geq 0$ and write

$$\begin{aligned} H_j(y) - H(y) &= \int \chi_{\{|f_j| > y\}} (\Phi(\nabla f_j) - \Phi(\nabla f)) d\mathcal{L}^n \\ (**) \quad &+ \int (\chi_{\{|f_j| > y\}} - \chi_{\{|f| > y\}}) (\Phi(\nabla f)) d\mathcal{L}^n. \end{aligned}$$

If this third conclusion were false, one could pass to a subsequence if necessary and assume $|H_j(y) - H(y)| > \gamma > 0$ for each j and pass to a further subsequence if necessary to insure that $\nabla f_j \rightarrow \nabla f$ almost everywhere and $f_j \rightarrow f$ almost everywhere as $j \rightarrow \infty$, and also to guarantee that there is some fixed summable function G with $|\nabla f| \leq G$ and $|\nabla f_j| \leq G$ for each j . It then follows that $\chi_{\{|f_j| > y\}} \rightarrow \chi_{\{|f| > y\}}$ almost everywhere on the complement of the set $\{x: |f(x)| = y\}$ on which $\nabla f = 0$ and $\Phi(\nabla f) = 0$ almost everywhere. The last integrand in (**) above is bounded by the function $|\Phi(\nabla f)|$ which is dominated by the summable function $|\nabla f|$; Lebesgue's dominated convergence theorem implies convergence of this last integral to zero. The continuity of Φ implies that the first integrand converges to zero almost everywhere. Since $\Phi(\nabla f_j) \leq |\nabla f_j| \leq G$, the first integrand is dominated by the summable function $2G$. The first integral thus also converges to zero. This concludes the proof under the extra hypothesis that $|\Phi(\xi)| \leq C|\xi|$. We now turn to the general case.

The function $\Phi(\xi) = |\xi|$ satisfies conditions (i) and (ii) above and our additional assumption as well. Hence, the three conclusions of our theorem hold for this integrand. We then use this fact to infer that for a general function Φ satisfying only conditions (i) and (ii), the three conclusions of our theorem hold for the function $\tilde{\Phi}(\xi) = |\xi| + |\Phi(\xi)|$ only if they hold for Φ itself. Therefore, we can replace Φ by $\tilde{\Phi}$ (and henceforth usually omit the tilde) and assume Φ satisfies the modified, weaker conditions

- (i)' $|\Phi(\xi)| \leq C|\Phi(\eta)|$ whenever $\xi, \eta \in \mathbf{R}^n$ with $|\xi| \leq 2|\eta|$;
- (ii) $|\Phi(\xi)| \leq C|\xi|$ for each $\xi \in \mathbf{R}^n$ with $|\xi| \leq 1$.

For the proof of conclusion (1), we use the same assumption (*) as before. Since $\Phi \circ \nabla f \in L^1(\mathbf{R}^n)$ and $\nabla f \in L^1(\mathbf{R}^n, \mathbf{R}^n)$, Lebesgue's monotone convergence theorem guarantees us the existence of a number T such that

$$C \int |\Phi(\nabla f)| \chi_{\{|\nabla f| > T\}} d\mathcal{L}^n \leq \frac{\alpha}{4}.$$

The conditions (i)' and (ii) on our new Φ also guarantee the existence of a constant C_T such that $|\Phi(\xi)| \leq C_T |\xi|$ whenever $|\xi| \leq 2T$. Let us write $\Phi = \hat{\Phi} + \Psi$ where $\hat{\Phi}(\xi) = \Phi(\xi)$ if $|\xi| \leq 2T$ and $\hat{\Phi}(\xi) = 0$ if $|\xi| > 2T$.

Since $\hat{\Phi}$ satisfies the extra hypothesis $|\hat{\Phi}(\xi)| \leq C_T |\xi|$, we can use our proof above to conclude that (1) holds for the sequence of functions

$$\hat{H}_{j,T}(y) = \int \chi_{\{|f_j| > y\}} \hat{\Phi}(\nabla f_j) d\mathcal{L}^n.$$

Therefore, we can concentrate on the remainder with Φ replaced by Ψ , namely,

$$H_{j,T}(y) = \int \chi_{\{|f_j| > y\}} \Psi(\nabla f_j) d\mathcal{L}^n \leq \int |\Psi(\nabla f_j)| d\mathcal{L}^n.$$

First consider the set $\{x : |\nabla f_j(x)| \geq 2|\nabla f(x)|\}$ in $\mathbf{R}^n$. On this set, $|\nabla f_j| \leq 2|\nabla f_j - \nabla f|$. From condition (i)', we conclude that

$$\int |\Psi(\nabla f_j)| \chi_{\{|\nabla f_j| \geq 2|\nabla f|\}} d\mathcal{L}^n \leq C \int |\Psi(\nabla f_j - \nabla f)| d\mathcal{L}^n,$$

and this last integral converges to zero as $j \rightarrow \infty$. For all large j 's, this term, in particular, will be less than $\frac{\alpha}{4}$. This means that it suffices to consider only the numbers

$$g_{j,T} = \int |\Psi(\nabla f_j)| \chi_{\{|\nabla f_j| < 2|\nabla f|\}} d\mathcal{L}^n$$

and to show that $g_{j,T} < \frac{\alpha}{4}$. However, $\Psi(\nabla f_j)$ vanishes unless $|\nabla f_j| > 2T$, and under the added condition that $|\nabla f_j| < 2|\nabla f|$, we have that $|\nabla f| > T$ when $\Psi(\nabla f_j) \neq 0$. Hence, by (i)' again, $|\Psi(\nabla f_j)| \leq C|\Phi(\nabla f)|$ and thus

$$g_{j,T} \leq C \int |\Phi(\nabla f)| \chi_{\{|\nabla f| > T\}} d\mathcal{L}^n \leq \frac{\alpha}{4}.$$

This concludes the proof of conclusion (1) in the general case.

For (2), we follow the earlier strategy and define L_j as before and $h_j = h_{j,L_j}$. Assuming the falsity of (2), we have

$$\beta < h_j \leq \int |\Phi(\nabla f_j)| L_j(f_j) d\mathcal{L}^n.$$

We again split $\mathbf{R}^n$ into the region $|\nabla f_j| \geq 2|\nabla f|$ and the region $|\nabla f_j| < 2|\nabla f|$. For the first region, we have $|\nabla f_j| \leq 2|\nabla f_j - \nabla f|$, and hence by (i)',

$$\begin{aligned} \int |\Phi(\nabla f_j)| \chi_{\{|\nabla f_j| \geq 2|\nabla f|\}} L_j(f_j) d\mathcal{L}^n &\leq C \int |\Phi(\nabla f_j - \nabla f)| L_j(f_j) d\mathcal{L}^n \\ &\leq C \int |\Phi(\nabla f_j - \nabla f)| d\mathcal{L}^n, \end{aligned}$$

but this converges to zero as $j \rightarrow \infty$. In the second region, $|\Phi(\nabla f_j)| < C|\Phi(\nabla f)|$ by condition (i)' since $|\nabla f_j| < 2|\nabla f|$. However,

$$\int |\Phi(\nabla f)| L_j(f_j) d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty$$

since $\Phi(\nabla f) \in L^1(\mathbf{R}^n, \mathbf{R}^n)$ and $L_j(f_j) \rightarrow 0$ almost everywhere. This establishes (2).

For (3), we pass to the necessary subsequences and proceed as before with the assumption $|H_j(y) - H(y)| > \gamma$ and note that we have to prove for each fixed γ that the right side of (**) converges to zero as $j \rightarrow \infty$. The second integral does so for the same reason as before because $\Phi \circ \nabla f \in L^1(\mathbf{R}^n)$. For the first integral, we still have that $\Phi(\nabla f_j) - \Phi(\nabla f) \rightarrow 0$ almost everywhere, but it remains to check that we can interchange integration and the limit $j \rightarrow \infty$. Again we split $\mathbf{R}^n$ into the region A where $|\nabla f_j| \geq 2|\nabla f|$ and the region where $|\nabla f_j| < 2|\nabla f|$. In the second region, as above we have $|\Phi(\nabla f_j)| \leq C|\Phi(\nabla f)|$ by condition (i)'. Thus, the integrand is dominated in the second region by the $L^1(\mathbf{R}^n)$ function $(C+1)|\Phi(\nabla f)|$, and the integral converges to zero as $j \rightarrow \infty$. In the first region, A , we set

$$X_j = |\Phi(\nabla f_j) - \Phi(\nabla f)| \quad \text{and} \quad Y_j = |\Phi(\nabla f_j - \nabla f)|,$$

and we have

$$X_j \leq Y_j + (X_j - Y_j)_+$$

so that

$$\int_A X_j d\mathcal{L}^n \leq \int_A Y_j d\mathcal{L}^n + \int_A (X_j - Y_j)_+ d\mathcal{L}^n.$$

Our hypotheses imply $\int Y_j d\mathcal{L}^n \rightarrow 0$ as $j \rightarrow \infty$. Also $(X_j - Y_j)_+ \rightarrow 0$ almost everywhere as $j \rightarrow \infty$. In A , we have $|\nabla f_j| \geq 2|\nabla f|$ which by condition (i)' implies

$$(X_j - Y_j)_+ \leq X_j \leq CY_j + |\Phi(\nabla f)|.$$

We have then $(X_j - Y_j)_+ \leq CY_j + ((X_j - Y_j)_+ - CY_j)_+$. But $\int Y_j d\mathcal{L}^n \rightarrow 0$ while the function $((X_j - Y_j)_+ - CY_j)_+$ converges to zero pointwise almost everywhere and is dominated by $|\Phi(\nabla f)|$ which is in $L^1(\mathbf{R}^n)$. This concludes the proof of (3). ■

3.4. Theorem (Convergence of cutoff $\mathcal{F}_j$'s). Suppose $f, f_1, f_2, \dots$ are functions from $\mathbf{R}^n$ to $\mathbf{R}^+$ which belong to $\mathbf{W}^{1,1}(\mathbf{R}^n)$, and suppose that $f_j \rightarrow f$ in $\mathbf{W}^{1,1}(\mathbf{R}^n)$ as $j \rightarrow \infty$. Fix $\varepsilon > 0$ and define $\Phi = \Phi_\varepsilon : \mathbf{R}^n \rightarrow \mathbf{R}^+$ by setting

$$\Phi_\varepsilon(\xi) = \frac{|\xi|^2}{\varepsilon^2 + |\xi|^2}.$$

Corresponding to our definition of co-area distribution functions $\mathcal{F}_f$, we define functions

$$(1) \quad \mathcal{F}_{(j)}(y) = \int \chi_{\{f_{(j)} > y\}} \Phi_\varepsilon(\nabla f_{(j)}) d\mathcal{L}^n,$$

where (j) denotes either j or no j . A trivial modification of Theorem 3.2 shows that $\mathcal{F}'_{(j)}$ belongs to $\mathbf{L}^1(\sigma, \infty)$ for each $\sigma > 0$. Our new assertion is that for every Borel subset A of (σ, ∞) having finite $\mathcal{L}^1$ measure, the functions $\mathcal{F}'_j$ converge to the function $\mathcal{F}'$ in measure. As is well known, this implies the existence of a subsequence $j(k)$ such that

$$(2) \quad \mathcal{F}'_{j(k)}(y) \rightarrow \mathcal{F}'(y)$$

as $k \rightarrow \infty$ for $\mathcal{L}^1$ almost every $y \in A$.

Proof. Suppose $\sigma > 0$ and that A is a Borel subset of $(2\sigma, \infty)$ with $\mathcal{L}^1(A) < \infty$. Suppose also $\beta > 0$ and set

$$A_j = A \cap \{y : \mathcal{F}'_j(y) - \mathcal{F}'(y) > \beta\} \quad \text{for each } j.$$

We need to show that $\mathcal{L}^1(A_j) \rightarrow 0$ as $j \rightarrow \infty$. (We also have to prove the same thing with “ $> \beta$ ” replaced by “ $< -\beta$,” but the proof is essentially the same.)

Suppose this is not true and (having passed to a subsequence if necessary) that $\mathcal{L}^1(A_j) > 5\alpha > 0$ for all j . Since our Φ satisfies the assumptions of Theorem 3.3, we know that the $\mathcal{F}'_j$'s are equi-absolutely continuous on $[\sigma, \infty)$. We can and do pick $\delta > 0$ sufficiently small so that the $\int_B |\mathcal{F}'_{(j)}| d\mathcal{L}^1 < \beta\alpha$ whenever B is a Borel subset of $[\sigma, \infty)$ with $\mathcal{L}^1(B) < \delta$. It is then straightforward to use Lusin's theorem to construct for each j a continuous function $L_j : \mathbf{R} \rightarrow [0, 1]$ having support in $[\sigma, \infty)$ such that the Borel set B_j on which L_j does not coincide with the characteristic function χ_{A_j} of A_j has measure less than δ . Hence, for each j ,

$$(3) \quad \int |L_j - \chi_{A_j}| |\mathcal{F}'_{(j)}| d\mathcal{L}^n \leq \int_{B_j} |\mathcal{F}'_{(j)}| d\mathcal{L}^n < \beta\alpha.$$

By our supposition,

$$\int \chi_{A_j} [\mathcal{F}'_j - \mathcal{F}'] d\mathcal{L}^n > \beta \mathcal{L}^1(A_j) > 5\beta\alpha.$$

On the other hand, by a trivial generalization of the formula of Theorem 3.2, we have

$$(4) \quad \int \chi_{A_j} \mathcal{F}'_{(j)} d\mathcal{L}^1 = - \int \chi_{A_j}(f_{(j)}) \Phi_\varepsilon(\nabla f_{(j)}) d\mathcal{L}^n.$$

We combine (3) and (4) to infer that for each j ,

$$I_j \equiv \int [L_j(f) \Phi_\varepsilon(\nabla f) - L_j(f_j) \Phi_\varepsilon(\nabla f_j)] d\mathcal{L}^n > 3\beta\alpha.$$

We set, as additional useful terminology, for each $y \in \mathbf{R}$ and each j ,

$$M_j(y) = \int_0^y L_j(s) d\mathcal{L}^1 s, \quad \rho_j = M_j \circ f_j, \quad \varphi_j = M_j \circ f$$

with distributional derivatives

$$\nabla \rho_j = L_j(f_j) \nabla f_j \quad \text{and} \quad \nabla \varphi_j = L_j(f) \nabla f.$$

We also define the vector-valued function

$$U_{(j)} = \frac{\nabla f_{(j)}}{\varepsilon^2 + |\nabla f_{(j)}|^2}.$$

Each I_j can then be written (with $A \cdot B = (A, B)$) as

$$I_j = \int [U \bullet \nabla \varphi_j - U_j \bullet \nabla \rho_j] d\mathcal{L}^n.$$

We then write for each j ,

$$I_j = \int U \bullet (\nabla \varphi_j - \nabla \rho_j) d\mathcal{L}^n + \int (U - U_j) \bullet \nabla \rho_j d\mathcal{L}^n.$$

Since $\nabla f_j \rightarrow \nabla f$ in $\mathbf{L}^1(\mathbf{R}^n, \mathbf{R}^n)$, there is a subsequence (which we continue to denote by j) and a function $G \in \mathbf{L}^1(\mathbf{R}^n)$ such that $|\nabla f_{(j)}| \leq G$ for each j and, as $j \rightarrow \infty$, $f_j \rightarrow f$ and $\nabla f_j \rightarrow \nabla f$ almost everywhere. From this it follows (since $|M'_j| = L_j \leq 1$) that $|\nabla \varphi_j| \leq G$ and $|\nabla \rho_j| \leq G$. Also, $U_j \rightarrow U$ almost everywhere as $j \rightarrow \infty$. Since $U_j \leq 1/\varepsilon$, we can use Lebesgue's dominated convergence theorem to infer that $\int (U - U_j) \bullet \nabla \rho_j d\mathcal{L}^n \rightarrow 0$ as $j \rightarrow \infty$. Therefore, for all sufficiently large j ,

$$(5) \quad \int U \bullet (\nabla \varphi_j - \nabla \rho_j) d\mathcal{L}^n > 2\beta\alpha.$$

The next step is to imitate the proof of Theorem 3.3 and replace U , which is in $\mathbf{L}^\infty(\mathbf{R}^n, \mathbf{R}^n) \cap \mathbf{L}^1(\mathbf{R}^n, \mathbf{R}^n)$, by a $\mathbf{C}_c^\infty(\mathbf{R}^n, \mathbf{R}^n)$ function W having the property that

$$(6) \quad \int [W - U] \bullet (\nabla \varphi_j - \nabla \rho_j) d\mathcal{L}^n < \beta\alpha$$

for all j . This can be done because of the bounds $|\nabla \varphi_j - \nabla \rho_j| \leq 2G \in \mathbf{L}^1(\mathbf{R}^n)$. Then from (5) and (6) and the definition of distributional derivative,

$$(7) \quad - \int (\varphi_j - \rho_j) \operatorname{div} W d\mathcal{L}^n = \int (\nabla \varphi_j - \nabla \rho_j) \bullet W d\mathcal{L}^n > \beta\alpha.$$

The fact that $|L_j| \leq 1$ implies that $|M_j(a) - M_j(b)| \leq |a - b|$. Since $f_j \rightarrow f$ almost everywhere, we also infer that $|\rho_j - \varphi_j| \rightarrow 0$ almost everywhere. Lebesgue's dominated convergence theorem implies that the left side of (7) converges to zero as $j \rightarrow \infty$. This gives our contradiction and proves the theorem. ■

Theorem 3.4 is the principal tool for proving the following main theorem of this section. When the cutoff ε is removed in (1) of Theorem 3.4, the continuity statement (2) of Theorem 3.4 is replaced by upper semicontinuity.

3.5. Theorem (Upper semicontinuity of the co-area distribution functions $\mathcal{F}_j$). Suppose $f, f_1, f_2, \dots$ are functions from $\mathbf{R}^n$ to $\mathbf{R}^+$ which belong to $\mathbf{W}^{1,1}(\mathbf{R}^n)$, and suppose that $f_j \rightarrow f$ in $\mathbf{W}^{1,1}(\mathbf{R}^n)$ as $j \rightarrow \infty$. We have the associated co-area distribution functions

$$(1) \quad \mathcal{F}_{(j)} = \mathcal{F}_{f_{(j)}}(y) = \int \chi_{\{f_{(j)} > y\}} \chi_{\{|\nabla f_{(j)}| \neq 0\}} d\mathcal{L}^n$$

where (j) denotes either j or no j . Then there is a subsequence $j(1), j(2), j(3), \dots$ of $1, 2, 3, \dots$ such that

$$\limsup_{k \rightarrow \infty} \mathcal{F}'_{j(k)}(y) \leq \mathcal{F}'(y)$$

for $\mathcal{L}^1$ almost every $y \geq 0$. Since the derivatives of the co-area distribution function are always nonpositive, we could equivalently conclude

$$\liminf_{k \rightarrow \infty} |\mathcal{F}'_{j(k)}(y)| \geq |\mathcal{F}'(y)|$$

for these y 's.

Proof. Suppose $\sigma > 0$ and $\mathcal{A}$ is a Borel subset of (σ, ∞) of finite measure. We will show for a subsequence $j(k)$ that $\limsup_{k \rightarrow \infty} \mathcal{F}'_{j(k)}(y) \leq \mathcal{F}'(y)$ for almost every $y \in \mathcal{A}$. In view of Cantor's diagonal process, this is sufficient to prove the theorem. For $\varepsilon = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$, we let Φ_ε be the function defined in Theorem 3.4 and use the notation $\mathcal{F}_{j_\varepsilon}$ and $\mathcal{F}_\varepsilon$ for the functions defined in (1) of Theorem 3.4. For each fixed value of our ε , we know the existence of a subsequence $j(k)$ such that $\mathcal{F}'_{j(k)\varepsilon}(y) \rightarrow \mathcal{F}'_\varepsilon(y)$ for $\mathcal{L}^1$ almost every $y \in \mathcal{A}$ as $k \rightarrow \infty$ in accordance with (2) of Theorem 3.4. Since we are considering only countably many ε 's, and the union of countably many sets of measure zero has measure zero, we can use Cantor's diagonal process to produce a single subsequence $j(l)$ of the j 's such that $\mathcal{F}'_{j(l)\varepsilon}(y) \rightarrow \mathcal{F}'_\varepsilon(y)$ for $\mathcal{L}^1$ almost every $y \in \mathcal{A}$ as $l \rightarrow \infty$ for each of our ε 's. We let $\mathcal{B}$ denote the set of y 's in $\mathcal{A}$ for which such simultaneous convergence occurs. Our sequence $j(l)$ will henceforth simply be denoted j .

We need to use four facts as follows:

- F1. $\mathcal{F}$ and the $\mathcal{F}_j$'s are absolutely continuous on (σ, ∞) .
- F2. $\lim_{j \rightarrow \infty} \mathcal{F}'_{j\varepsilon}(y) = \mathcal{F}'_\varepsilon(y)$ for each of our ε 's and each $y \in \mathcal{B}$.
- F3. $\lim_{\varepsilon \rightarrow 0} \mathcal{F}_{(j)\varepsilon}(y) = \mathcal{F}_{(j)}(y)$ for each $y > 0$ and all (j) .
- F4. $\mathcal{F}'_{(j)}(y) \leq \mathcal{F}'_{(j)\delta}(y) \leq \mathcal{F}'_{(j)\varepsilon}(y)$ for $\mathcal{L}^1$ almost every y and $\delta < \varepsilon$ and all (j) .

F1 follows from Theorem 3.2. F2 follows from Theorem 3.4. F3 is a consequence of the definitions and Lebesgue's dominated convergence theorem. F4 is evident from the identity (4) in the proof of Theorem 3.4; in particular, for every Borel set $A \subset (\sigma, \infty)$,

$$\int_A \mathcal{F}'_\varepsilon d\mathcal{L}^1 = - \int \chi_{\{f \in A\}} \frac{|\nabla f|^2}{|\nabla f|^2 + \varepsilon^2} d\mathcal{L}^n,$$

with corresponding formulas for the f_j 's and $\mathcal{F}_j$'s.

Our first new assertion is that for each (j) ,

$$(2) \quad \lim_{\varepsilon \rightarrow 0} \mathcal{F}'_{(j)\varepsilon}(y) = \mathcal{F}'_{(j)}(y)$$

for $\mathcal{L}^1$ almost every y in $\mathcal{B}$. To prove this, write

$$\begin{aligned} \mathcal{F}_{(j)}(y) &= - \int_y^\infty \mathcal{F}'_{(j)} d\mathcal{L}^1 \\ &= \lim_{\varepsilon \rightarrow 0} \mathcal{F}_{(j)\varepsilon}(y) \quad (\text{by F3}) \\ &= - \lim_{\varepsilon \rightarrow 0} \int_y^\infty \mathcal{F}'_{(j)\varepsilon} d\mathcal{L}^1 \\ &= - \int_y^\infty \lim_{\varepsilon \rightarrow 0} \mathcal{F}'_{(j)\varepsilon} d\mathcal{L}^1 \quad (\text{by F4 and monotone convergence}). \end{aligned}$$

This implies that $\int_y^\infty (\mathcal{F}'_{(j)} - \lim_{\varepsilon \rightarrow 0} \mathcal{F}'_{(j)\varepsilon}) d\mathcal{L}^1 = 0$, which, together with F4, implies (2). Since, again, the countable union of sets of measure zero has measure zero, we include the existence of a subset $\mathcal{B}_0$ of $\mathcal{B}$ on which (2) holds for each (j) and each $y \in \mathcal{B}_0$ and for which $\mathcal{L}^1(\mathcal{A} \sim \mathcal{B}_0) = 0$.

From (2) and F2, we conclude that for every $y \in \mathcal{B}_0$,

$$(3) \quad \mathcal{F}'(y) = \lim_{\varepsilon \rightarrow 0} \mathcal{F}'_\varepsilon(y) = \lim_{\varepsilon \rightarrow 0} \lim_{j \rightarrow \infty} \mathcal{F}'_{j\varepsilon}(y).$$

On the other hand, by F4, $\mathcal{F}'_{j\varepsilon}(y) \geq \mathcal{F}'_j(y)$ for all $y \in \mathcal{B}_0$ so that

$$(4) \quad \lim_{j \rightarrow \infty} \mathcal{F}'_{j\varepsilon}(y) \geq \limsup_{j \rightarrow \infty} \mathcal{F}'_j(y).$$

Combining (4) with (3), one obtains the desired conclusion. ■

3.6. The co-area formula and co-area regularity. The co-area formula of Federer [FH, 3.2.11] as extended by Brothers and Ziemer is motivation not only for the appellation “co-area regularity” but it is also the tool we originally used (together with machinery from geometric measure theory) to analyze the problem of the continuity of symmetric decreasing rearrangement. (This original approach is summarized in a conference report [AL].) Theorem 3.5, discovered subsequently, eliminates the need for this machinery, but we record here the basic ideas and connection with Theorem 3.5.

The co-area formula implies that if f belongs to our class $\mathcal{C}_2$ and if $g : \mathbf{R}^n \rightarrow \mathbf{R}^+$ is a summable Borel function, then

$$(1) \quad A(y) \equiv \int_{f^{-1}\{y\}} g d\mathcal{H}^{n-1}$$

exists for $\mathcal{L}^1$ almost every $y \in \mathbf{R}^+$; here $\mathcal{H}^{n-1}$ is Hausdorff's $(n-1)$ -dimensional measure over $\mathbf{R}^n$. Moreover, A belongs to $\mathbf{L}^1(\mathbf{R})$ and

$$(2) \quad \int A d\mathcal{L}^1 = \int g |\nabla f| d\mathcal{L}^n.$$

Now fix $Y > 0$, and let us make the choice $g = g_{\delta,Y} \equiv \chi_{\{f>Y\}}(|\nabla f| + \delta)^{-1}$ for positive numbers δ . As $\delta \rightarrow 0$, a simple monotone convergence argument based on (1) shows that

$$(3) \quad h_Y(y) \equiv \begin{cases} \int_{f^{-1}\{y\}} |\nabla f|^{-1} d\mathcal{H}^{n-1} & \text{for } y > Y, \\ 0 & \text{for } y \leq Y \end{cases}$$

is finite for $\mathcal{L}^1$ almost every y so that (2) then implies

$$(4) \quad \int h_Y d\mathcal{L}^1 = \int \chi_{\{f>y\}} \chi_{\{|\nabla f| \neq 0\}} d\mathcal{L}^n = \mathcal{F}_f(Y).$$

This equation shows that $\mathcal{F}_j$ is absolutely continuous and that for $\mathcal{L}^1$ almost every y ,

$$(5) \quad \mathcal{F}'_f(y) = - \int_{f^{-1}\{y\}} |\nabla f|^{-1} d\mathcal{H}^{n-1}.$$

Theorem 3.2 shows by a completely different method that $\mathcal{F}_f$ is absolutely continuous, but it does not display $\mathcal{F}'_f$ explicitly as in (5); we do not need this explicit formula.

We are interested in $\mathcal{V}_f = \mathcal{F}_f + \mathcal{G}_f$, and, at first, one is inclined to think, from a hasty perusal of the co-area formula, that $\mathcal{V}'_f$ equals the right side of (5). This is false. The co-area formula says quite a bit about $\mathcal{F}'_f$, but *the co-area formula is blind to $\mathcal{G}'_f$* .

Theorem 3.5 tells us that if $f_j \rightarrow f$, then a subsequence of $\mathcal{F}'_{f_j}$ converges to $\mathcal{F}'$ upper semicontinuously (recall that these derivatives are all nonpositive). In our original approach, this theorem was difficult for it required the machinery of rectifiable currents and of slicing. Theorem 3.5 simplifies this considerably. We shall explain the geometric measure theory approach briefly below, but for the moment let us note the relation to co-area regularity.

The *key fact* is that when f is co-area regular, the absolutely continuous part of $\mathcal{V}'_f$ comes entirely from $\mathcal{F}'_f$. If now $f_j \rightarrow f$, it may happen that the $\mathcal{G}'_{f_j}$'s have absolutely continuous parts (which would necessarily be negative). If so, that only strengthens the upper semicontinuity result for the absolutely continuous part $(\mathcal{V}'_f)^{\text{a.c.}}$, namely,

$$(6) \quad \limsup_{j \rightarrow \infty} (\mathcal{V}'_{f_j})^{\text{a.c.}} \leq (\mathcal{V}'_f)^{\text{a.c.}} = \mathcal{F}'_f.$$

Inequality (6) is what is needed in §9 to obtain arc length convergence and, ultimately, convergence of $\nabla f_j^* \rightarrow \nabla f^*$. Geometric measure theory is not needed for this last step.

Finally, let us outline the geometric measure theory approach to Theorem 3.5. These methods are in part adaptations of and extensions of the methods of [AF] and [TA]. We use the Euclidean currents $\mathbf{E}^{n+1} = \mathbf{E}^n \times \mathbf{E}^1$ to construct an n -dimensional locally rectifiable current $Q = \mathbf{E}^{n+1} \llcorner \{(x, y) : y < f(x)\}$ from the subgraph of f . The boundary n -dimensional locally rectifiable current

$T = \partial Q$ is associated with the graph of f . Similar constructions produce Q_j 's and T_j 's associated with the f_j 's. The assumed $L^1(\mathbf{R}^n)$ convergence of the f_j 's to f implies that the masses of the currents $Q_j - Q$ converge to zero. We let $\pi: \mathbf{R}^n \times \mathbf{R} \rightarrow \mathbf{R}$ and $\Pi: \mathbf{R}^n \times \mathbf{R} \rightarrow \mathbf{R}^n$ denote projections onto the corresponding factors. These mappings π and Π induce mappings $\pi_\#$ and $\Pi_\#$ on spaces of currents which commute with the boundary operator.

A reasonable real-valued function g defined on a surface S generally decomposes S into a union of $(n-1)$ -dimensional level sets $S \cap g^{-1}\{y\}$, $y \in \mathbf{R}$. When S also carries the structure of a rectifiable current, the theory of slicing set forth in [FH, 4.3] shows how to give a corresponding current structure to these level sets; the current associated to the level set $S \cap g^{-1}\{y\}$ is denoted $\langle S, g, y \rangle$ and is called the *slice of S by g at y* . According to this theory of slicing, we know for each j and almost every $y > 0$ that all the slices below exist and satisfy the conditions

$$\begin{aligned}\Pi_\# \langle T, \pi, y \rangle &= (-1)^n \langle \mathbf{E}^n, f, y \rangle, \\ \Pi_\# \langle T_j, \pi, y \rangle &= (-1)^n \langle \mathbf{E}^n, f_j, y \rangle, \\ \langle T_j - T, \pi, y \rangle &= -\partial \langle Q_j - Q, \pi, y \rangle.\end{aligned}$$

The estimate that the masses of the $Q_j - Q$ converge to zero implies (for subsequences) that the currents $\langle Q_j - Q, \pi, y \rangle$ also converge to zero for almost every y . This, in turn, implies that the currents $\Pi_\# \langle T_j, \pi, y \rangle$ converge weakly to the current $\Pi_\# \langle T, \pi, y \rangle$ for almost every y . The lower semicontinuity of mass under weak convergence of currents effectively translates into the upper semicontinuity of convergence of the derivatives $\mathcal{F}'_{f_j}$ indicated above. Proof by this route (which we followed initially) is quite technical.

§4. THE MAIN THEOREMS

As we said in the introduction, some conditions are needed in order to insure the convergence of the gradients of symmetric decreasing rearrangements. This section contains our main results on this question. Three theorems are stated. The first, Theorem 4.1, states some very general conditions under which convergence is assured. The important point here is (3). Theorem 4.2 utilizes (3) to show, with the help of results of §3, that co-area regularity guarantees convergence. Theorem 4.3, which is a converse to Theorem 4.2, states that it is possible to construct sequences for which the gradients of rearrangements do not converge. Theorem 4.1 is proved in 8.8. Theorem 4.2 is proved in 8.9. Theorem 4.3 is proved in 6.1.

4.1. Theorem (Derivatives of distribution functions and the continuity of rearrangement). *We assume as general hypotheses that*

(a) $f, f_1, f_2, f_3, \dots$ are functions mapping $\mathbf{R}^n$ to $\mathbf{R}^+$ which belong to our class $\mathcal{C}_1$ (see Definition 1.2.2);

(b) $f_j \rightarrow f$ in measure as $j \rightarrow \infty$;

(c) $\Psi: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is a suitable integrand (see Definition 2.1) such that $\Psi(|\nabla f|)$ and the $\Psi(|\nabla f_j|)$'s are $\mathcal{L}^n$ summable;

- (d) $\int \Psi(|\nabla f_j - \nabla f|) d\mathcal{L}^n \rightarrow 0$ as $j \rightarrow \infty$;
- (e) $f^*, f_1^*, f_2^*, f_3^*, \dots$ are the symmetric decreasing rearrangements of the functions in (a) (see Definition 1.3);
- (f) $\mathcal{V}_f, \mathcal{V}_{f_1}, \mathcal{V}_{f_2}, \mathcal{V}_{f_3}, \dots : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ are the monotonically nonincreasing distribution functions (see Definition 1.2.5) associated with the functions in (a) (equivalently, of the functions in (e));
- (g) $\mathcal{W}, \mathcal{W}_1, \mathcal{W}_2, \mathcal{W}_3, \dots : \mathbf{R}^+ \rightarrow \mathbf{R}^-$ are the classical derivatives of the distribution functions in (f), e.g., the distribution first derivative of $\mathcal{V}_f$ can be written $\mathcal{L}^1 \wedge \mathcal{W} - \nu$ for some Radon measure ν over $\mathbf{R}^+$ which is singular with respect to Lebesgue's measure, etc. (see Definition 1.2.5). (Note that $\mathcal{W}$ was called $(\mathcal{V}_f')^{a.c.}$ in equation (6) of 3.6, etc.)

Then

$$(1) \quad \int |\Psi(|\nabla f_j|) - \Psi(|\nabla f|)| d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

Furthermore, the following two conditions are equivalent.

$$(2) \quad \int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

(3) Each subsequence of the j 's has a further subsequence $j(1), j(2), j(3), \dots$ such that

$$\limsup_{k \rightarrow \infty} \mathcal{W}_{j(k)}(y) \leq \mathcal{W}(y) \quad \left(\text{equivalently, } \liminf_{k \rightarrow \infty} |\mathcal{W}_{j(k)}(y)| \geq |\mathcal{W}(y)| \right)$$

for $\mathcal{L}^1$ almost every $0 < y < \infty$.

Each of these conditions implies further

$$(4) \quad \int |\Psi(|\nabla f_j^*|) - \Psi(|\nabla f^*|)| d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

Remark. The main hypothesis of this theorem is equation (d) about the convergence of the gradients. However, some amount of control of $|f_j(x) - f(x)|$ is also needed; the principal reason for this is that functions in $\mathcal{E}_1$ (unlike functions in $\mathcal{E}_2$) can have very long tails as $|x| \rightarrow \infty$. In this theorem, the control is achieved by the fairly weak assumption of convergence of the f_j 's to f in measure (i.e., for each $\varepsilon > 0$, we have $\mathcal{L}^n\{x : |f_j(x) - f(x)| > \varepsilon\} < \varepsilon$ for all j 's which are sufficiently large). Theorem 2.4 guarantees that the f_j^* 's also converge in measure to f^* . Such convergence insures that every subsequence of the f_j 's contains a further subsequence with the property $f_j(x) \rightarrow f(x)$ pointwise $\mathcal{L}^n$ almost everywhere. On the other hand, convergence in measure is assured by the more commonly adopted hypothesis that $f_j \rightarrow f$ in some $L^p(\mathbf{R}^n)$ space. Indeed, convergence in measure is assured by the condition $\int \Phi(|f_j - f|) d\mathcal{L}^n \rightarrow 0$ for any integrand $\Phi : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ with the property that for each $\varepsilon > 0$ there is some R_ε such that $\Phi(x) > \varepsilon$ whenever $|x| > R_\varepsilon$.

4.2. Theorem (Rearrangement is continuous at co-area regular functions). Suppose that $f, f_1, f_2, f_3, \dots$ are functions mapping $\mathbf{R}^n$ to $\mathbf{R}^+$ which belong

to our class $\mathcal{E}_1$ (see Definition 1.2.2). Let $f^*, f_1^*, f_2^*, f_3^*, \dots$ denote the symmetric decreasing rearrangements of these functions (see Definition 1.3). Assume that $f_j \rightarrow f$ in measure as $j \rightarrow \infty$. Suppose also that $\Psi: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is a suitable integrand (see Definition 2.1) such that $\Psi(|\nabla f|)$ and the $\Psi(|\nabla f_j|)$'s are $\mathcal{L}^n$ summable. Finally, suppose that

$$(1) \quad \int \Psi(|\nabla f_j - \nabla f|) d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

Then

$$(2) \quad \int |\Psi(|\nabla f_j|) - \Psi(|\nabla f|)| d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

In case our limit function f is co-area regular (see Definition 1.2.6), then the following additional convergences are guaranteed:

$$(3) \quad \int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty,$$

$$(4) \quad \int |\Psi(|\nabla f_j^*|) - \Psi(|\nabla f^*|)| d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

4.3. Theorem (Rearrangement is not continuous at co-area irregular functions). Suppose that f is a co-area irregular function mapping $\mathbf{R}^n$ to $\mathbf{R}^+$ which belongs to our class $\mathcal{E}_1$ (see Definition 1.2.2). By definition, this co-area irregularity condition means that the distribution first derivative of the residual distribution function

$$\mathcal{E}_f(y) = \int \chi_{\{f>y\}} \chi_{\{\nabla f=0\}} d\mathcal{L}^n$$

(when regarded as a measure on $\mathbf{R}^+$) has an absolutely continuous component with respect to Lebesgue measure $\mathcal{L}^1$. This means, in particular, that there is a nonzero $L^1_{\text{loc}}(\mathbf{R}^+)$ function $h: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ such that for each Borel subset A of $\mathbf{R}^+$,

$$\int \chi_{\{f \in A\}} \chi_{\{\nabla f=0\}} d\mathcal{L}^n \geq \int_A h d\mathcal{L}^1.$$

Now fix any numbers $0 < Y < Z < \infty$ for which

$$\int_{Y+(Z-Y)/4}^{Z-(Z-Y)/4} h d\mathcal{L}^1 > 0.$$

Then there is a sequence $f_1, f_2, f_3, \dots$ of functions mapping $\mathbf{R}^n$ to $\mathbf{R}^+$ which belongs to our class $\mathcal{E}_1$ such that ∇f_j^* does not converge to ∇f^* in $L^1_{\text{loc}}(\mathbf{R}^n, \mathbf{R}^n)$, but for which the f_j 's converge to f in the following senses.

(1) $f_j - f \rightarrow 0$ in $L^\infty(\mathbf{R}^n)$ with

$$f_j(x) = f(x) \quad \text{whenever } 0 < f(x) \leq Y \text{ or } f(x) \geq Z$$

and

$$Y < f_j(x) < Z \quad \text{whenever } Y < f(x) < Z$$

for each j .

(2) $\nabla f_j - \nabla f \rightarrow 0$ in $\mathbf{L}^1(\mathbf{R}^n, \mathbf{R}^n)$, and there is a constant C such that $|\nabla f_j(x)| \leq C|\nabla f(x)| + 1$ for $\mathcal{L}^n$ almost every x .

(3) The $\mathcal{L}^n$ measure of the set of x 's in $\mathbf{R}^n$ for which $\nabla f(x) \neq 0$ and for which $\nabla f_j(x) \neq \nabla f$ converges to zero as $j \rightarrow \infty$.

§5. EXISTENCE OF CO-AREA IRREGULAR FUNCTIONS

5.1. Construction of a $C^{n-1,\lambda}(\mathbf{R}^n)$ co-area irregular function. Corresponding to each $n \geq 2$, each $0 < \lambda < 1$, and each $0 < \alpha < 1$, we here construct a function $f: \mathbf{R}^n \rightarrow \mathbf{R}^+$ with the following properties.

(1) The support of f is the central n cube of side length 2 and volume 2^n ,

$$Q_0 = \mathbf{R}^n \cap \{x : |x_i| \leq 1 \text{ for } i = 1, \dots, n\}.$$

(2) The range of f is the unit interval $[0, 1]$.

(3) All partial derivatives of f of order less than n are continuous, and the partial derivatives of f of order $n-1$ are Hölder continuous with exponent λ . We abbreviate these conditions, as usual, by saying that f belongs to the function space $C^{n-1,\lambda}(\mathbf{R}^n)$.

(4) The residual distribution function $\mathcal{G}_f$ of f (see Definition 1.2.5) is purely absolutely continuous with respect to Lebesgue's measure $\mathcal{L}^1$ on $\mathbf{R}^+$. More precisely, there is a positive constant $C = 2^n(1-\alpha)$ for which $\mathcal{G}_f(y) = C(1-y)$ for $0 \leq y \leq 1$ while, of course, $\mathcal{G}_f(y) = 0$ for $y \geq 1$.

(5) $\mathcal{L}^n(Q_0 \cap \{x : \nabla f = 0\}) = (1-\alpha)\mathcal{L}^n(Q_0) = (1-\alpha)2^n$

Thus, our f maximally contradicts any preconception that $\mathcal{G}'_f(y)$ must be a singular measure. The differentiability in (3) is also optimal because, as we show in Theorem 5.2 below, if $f \in C^{n-1,1}(\mathbf{R}^n)$, then $\mathcal{G}'_f(y)$ must be purely singular. Theorem 5.4 states that when $n = 1$, all f 's in $\mathcal{E}_1$ are co-area regular.

Our construction is an adaptation of Whitney's construction [WH] of a smooth function which is not constant on a connected component of its critical set. The reason that our proof is much longer than Whitney's is that we wish to have more control over the details of the function (notably properties (3), (4), (5)) than does Whitney. Another difference is that we use a triadic decomposition of the cube instead of a dyadic decomposition as Whitney does. For the purposes of 5.1, a dyadic decomposition could be equally well employed. It turns out, however, that the triadic function we construct has additional properties which, while not explicitly mentioned above, will be important for us in Theorem 6.2, where we show the following. If f is as above and if $g: \mathbf{R}^n \rightarrow \mathbf{R}^+$ is a smooth function with compact support whose restriction to Q_0 is a suitable affine function, then there are well-chosen positive numbers $t_1, t_2, t_3, \dots$ converging to 0 such that $f + t_j g$ converges to f in $\mathbf{W}^{1,1}(\mathbf{R}^n)$ (obviously), but the symmetrized functions $(f + t_j g)^*$ do not converge to f^* in $\mathbf{W}^{1,1}(\mathbf{R}^n)$ as $j \rightarrow \infty$.

Let us begin with some heuristic and imprecise remarks which might help to clarify our construction of f . Our aim is to construct an n -dimensional Cantor-type set (which will also be the Cartesian product of 1-dimensional Cantor-type sets). First divide Q_0 into 3^n cubes of one-third the size. Index these cubes by labels $(0), (1), \dots, (3^n - 1)$ in any convenient way. Now let $f_1: \mathbf{R}^n \rightarrow [0, 1]$ map everything outside Q_0 to zero, and take value $3^{-n}m$ in cube (m) for each m . Next divide each of these first generation cubes (m) into 3^n cubes of one-third the size (i.e., the edge length of each of the second generation cubes is one-ninth the edge length of Q_0), and index these cubes by labels (m, k) corresponding to each $m = 0, \dots, 3^n - 1$ and each $k = 0, \dots, 3^n - 1$. Let $f_2: \mathbf{R}^n \rightarrow [0, 1]$ map everything outside Q_0 to zero, and take value $3^{-2n}k$ in cube (m, k) for each m and k . Continue this subdivision process, and define functions f_3, f_4 , etc. In particular, in the j th generation cube indexed by $(m, k, \dots, l)$, the function f_j takes value $3^{-jn}l$. We then set $f = \sum_{j=1}^{\infty} f_j$. Clearly, the range of f is the entire closed interval $[0, 1]$. This f is not continuous in any nonempty open subset of Q_0 , much less differentiable. It is, however, a model for the real f we wish to construct. To achieve continuity and, indeed, differentiability up to order $n - 1$, we first replace f_1 by an infinitely differentiable function which, corresponding to each cube (m) , assumes the value $3^{-n}m$ in an inner cube which is smaller than cube (m) but which is concentric with it; this smaller cube will have side length equal to $(\frac{2}{3})(1 - \delta\gamma)$ for a well-chosen number $0 < \gamma < 1$ and for a number δ with $0 < \delta < 1$. γ is determined by λ while δ is determined by α . Thus, the cube (m) now has an inner cube of side length $(\frac{2}{3})(1 - \delta\gamma)$ on which f_1 is constant. Similarly, f_2 is modified to be infinitely differentiable, but its inner cube has side length $\frac{2}{9}(1 - \delta\gamma)(1 - \delta\gamma^2)$, instead of $\frac{2}{9}(1 - \delta\gamma)(1 - \delta\gamma)$. Moreover, the center of the (m, k) cube is moved a bit so that its center coincides with the center of the 3^n cubes formed by dividing the *inner* part of cube (m) into 3^n pieces. The process is repeated indefinitely as before (with a factor $(1 - \delta\gamma^k)$ at the k th generation). It turns out that a reasonable interpolation of f_j is made in the regions between these various inner cubes, and if, as before, we set $f = \sum_j f_j$, then $f \in C^{n-1}(\mathbf{R}^n)$ provided $3^{-1/(n-1)} < \gamma < 1$ and $f \in C^{n-1, \lambda}(\mathbf{R}^n)$ if $2^{\lambda-1} \leq \gamma^{\lambda+n-1} < 1$. More importantly, ∇f will be zero on the intersection of all the inner cubes, which is a set of positive measure. As was the case before, f assumes each value in the interval $[0, 1]$ on the intersection of the inner cubes, and it turns out that $\mathcal{E}'_f(y) \neq 0$ for *all* values of y (not just a set of measure zero).

We turn to the details of the construction.

Step 1. Choose $\gamma \in (3^{-1/(n-1)}, 1)$ and $0 < \delta < 1$, and define lengths L_j for $j = 1, 2, \dots$ which will be the half-lengths of our (outer) cubes by setting $L_1 = \frac{1}{3}$ and

$$L_j = 3^{-j} \prod_{m=1}^{j-1} (1 - \delta\gamma^m) \quad \text{for } j = 2, 3, \dots$$

We denote by Σ the collection of all infinite sequences

$$\sigma = \{\sigma(1), \sigma(2), \sigma(3), \dots\}$$

with each $\sigma(j)$ in $\mathbf{R}^n$ of the form

$$\sigma(j) = (\sigma(j)_1, \sigma(j)_2, \dots, \sigma(j)_n) \in \{-1, 0, +1\}^n \subset \mathbf{R}^n.$$

Associated with each sequence $\sigma \in \Sigma$ and each positive integer j is the integer

$$\begin{aligned} N_j(\sigma) &= \left[\sigma(j)_1 + 3\sigma(j)_2 + \dots + 3^{n-1}\sigma(j)_n + \frac{3^n - 1}{2} \right] \\ &= \sum_{s=1}^n (\sigma(j)_s + 1)3^{s-1} \in \{0, 1, \dots, 3^n - 1\} \end{aligned}$$

which will be used in defining our final function. The value of $N_j(\sigma)$ depends only on $\sigma(j)$, the j th element of the sequence, and as the $\sigma(j)$'s vary over the 3^n possible values in $\{-1, 0, +1\}^n$, the function $N_j(\sigma)$ assumes each of the values $0, 1, 2, \dots, 3^n - 1$ exactly once.

Associated with each sequence $\sigma \in \Sigma$ and each positive integer J is the center point at the J th generation given by

$$C_J(\sigma) = 2 \sum_{j=1}^J L_j \sigma(j) \in Q_0.$$

Further associated with the same σ and J is the open (outer) cube

$$Q_J(\sigma) = \{x : \|x - C_J(\sigma)\| < L_J\} \subset Q_0.$$

Here the double bar $\|\cdot\|$ denotes the $\mathbf{L}^\infty$ norm (not the Euclidean norm) in $\mathbf{R}^n$ given by setting

$$\|x\| = \max_i |x_i|$$

for each x . Clearly, for each σ and J , we have

$$Q_J(\sigma) \subset Q_{J-1}(\sigma) \subset \dots \subset Q_1(\sigma) \subset Q_0.$$

Additionally, associated with the same σ and J is the open (inner) cube

$$P_J(\sigma) = \{x : \|x - C_J(\sigma)\| < L_J(1 - \delta\gamma^J)\} \subset \{x : \|x - C_J(\sigma)\| < L_J\} = Q_J(\sigma).$$

Clearly, $C_J(\sigma)$, $Q_J(\sigma)$, and $P_J(\sigma)$ depend only on $\sigma(1), \dots, \sigma(J)$, e.g., $Q_J(\sigma) = Q_J(\tilde{\sigma})$ in case $\sigma(j) = \tilde{\sigma}(j)$ for each $j = 1, \dots, J$.

We make the following three assertions (A), (B), and (C).

(A) Suppose that σ and $\tilde{\sigma}$ belong to Σ with $\sigma(K) \neq \tilde{\sigma}(K)$. Then $Q_J(\sigma) \cap Q_J(\tilde{\sigma})$ is empty for each $J \geq K$. Thus, for every σ and $\tilde{\sigma}$, there is the dichotomy that either $Q_J(\sigma) = Q_J(\tilde{\sigma})$ or else $Q_J(\sigma)$ and $Q_J(\tilde{\sigma})$ are disjoint.

Proof. We suppose without loss of generality that K is the smallest number with the property that $\sigma(K) \neq \tilde{\sigma}(K)$ and that $J \geq K$. If our assertion were false, there would exist some x such that

$$\left\| x - 2 \sum_{j=1}^J \sigma(j) L_j \right\| < L_J \quad \text{and} \quad \left\| x - 2 \sum_{j=1}^J \tilde{\sigma}(j) L_j \right\| < L_J.$$

The triangle inequality then implies

$$2L_J > \left\| 2 \sum_{j=1}^J [\sigma(j) - \tilde{\sigma}(j)] L_j \right\| = \left\| 2 \sum_{j=K}^J [\sigma(j) - \tilde{\sigma}(j)] L_j \right\|.$$

Since $\|\sigma(j) - \tilde{\sigma}(j)\| \in \{-2, -1, 0, 1, 2\}$, the final expression can be estimated as follows.

If $J = K$, then the value is at least $2L_J$, which clearly contradicts the inequality.

If $J > K$, then the value is not less than $2L_K - 4 \sum_{j=K+1}^J L_j$.

We note that when $j \geq K$, then $L_j \leq 3^{K-j} L_K$ so that

$$\sum_{j=K+1}^J L_j < \frac{1}{2} L_K (1 - 3^{K-J})$$

which would imply that $L_J > 3^{K-J} L_K$, which again is false. ■

(B) Suppose that σ and $\tilde{\sigma}$ belong to Σ . If $\sigma(j) = \tilde{\sigma}(j)$ for each $j = 1, \dots, K$, then $Q_J(\sigma) \subset P_K(\tilde{\sigma})$ for each $J > K$. If $\sigma(j) \neq \tilde{\sigma}(j)$ for some $j \in \{1, \dots, K\}$, then $Q_J(\sigma) \cap Q_K(\tilde{\sigma})$ is empty for each $J > K$.

Proof. We first show that the second assertion of (B) is a consequence of the first assertion of (B) and assertion (A). By (A), we have $Q_K(\tilde{\sigma}) \cap Q_K(\sigma)$ is empty. By the first assertion in (B), we have (setting $\tilde{\sigma} = \sigma$) that $Q_J(\sigma) \subset P_K(\sigma)$. But $P_K(\sigma) \subset Q_J(\sigma)$ as noted earlier, and thus $P_J(\sigma) \cap Q_K(\sigma)$ is empty.

Turning now to the first assertion of (B), we suppose that $\sigma(j) = \tilde{\sigma}(j)$ for each $j = 1, \dots, K$, that $J > K$, and that

$$\left\| x - 2 \sum_{j=1}^J \sigma(j) L_j \right\| < L_J.$$

We must show that

$$\left\| x - 2 \sum_{j=1}^K \sigma(j) L_j \right\| < L_K (1 - \delta \gamma^K).$$

Our assumed inequality implies

$$L_J > \left\| x - 2 \sum_{j=1}^K \sigma(j) L_j \right\| - 2 \sum_{j=K+1}^J L_j.$$

If $j > K$, then $L_j \leq 3^{K-j} (1 - \delta \gamma^K) L_K$ which implies

$$2 \sum_{j=K+1}^J L_j \leq (1 - 3^{K-J}) (1 - \delta \gamma^K) L_K \leq (1 - \delta \gamma^K) L_K - L_J,$$

which is readily checked to be the required inequality. ■

(C) Suppose $\sigma \in \Sigma$ and J is a positive integer. Then

$$\text{Clos}[P_J(\sigma)] = \bigcup \{ \text{Clos}[Q_{J+1}(\tilde{\sigma})] : \sigma(j) = \tilde{\sigma}(j) \text{ for each } j = 1, \dots, J \}.$$

Proof. Trivial. ■

Step 2. For each $0 < \varepsilon < 1$, standard constructions produce infinitely differentiable, symmetric functions $g_\varepsilon, h_\varepsilon : \mathbf{R} \rightarrow [0, 1]$ with the following properties:

- (a) $g_\varepsilon(t) = h_\varepsilon(t) = 0$ if $|t| \geq 1$.
- (b) $g_\varepsilon(t) = 0$ and $h_\varepsilon(t) = 1$ if $|t| \leq 1 - \varepsilon$, and, for all t , $0 \leq g_\varepsilon(t) \leq 1$ and $0 \leq h_\varepsilon(t) \leq 1$.
- (c) $g'_\varepsilon(t) > 0$ if $1 - \varepsilon < t < 1 - \frac{\varepsilon}{2}$ and $g'_\varepsilon(t) < 0$ if $1 - \frac{\varepsilon}{2} < t < 1$.
- (d) $h'_\varepsilon(t) < 0$ if $1 - \varepsilon < t < 1$.
- (e) There exists a constant $A > 1$ which is independent of ε such that $|(d^k g_\varepsilon/dt^k)(s)| < A/\varepsilon^k$ and $|(d^k h_\varepsilon/dt^k)(s)| < A/\varepsilon^k$ for each real number s and each positive integer $k \leq n - 1$.

For each positive integer j , we set $G_j = g_{\delta\gamma^j}$ and $H_j = h_{\delta\gamma^j}$, i.e., ε is set equal to $\delta\gamma^j$. We now define functions $M_j^0, M_j : \mathbf{R}^n \rightarrow [0, 1]$ for each positive integer j by setting for each x

$$M_j^0(x) = \prod_{i=1}^n G_j(x_i)$$

and

$$M_j(x) = \prod_{i=1}^n H_j(x_i).$$

Using these functions M_j^0 and M_j , we shall now define our function f which will satisfy conditions (1)–(5). There are three cases as follows.

Case 0. We denote by S_0 the set of x 's in $\mathbf{R}^n$ which do not belong to any of the $Q_j(\sigma)$'s for any positive integer j and any $\sigma \in \Sigma$ and define $f(x) = 0$ for such x 's. The set S_0 consists of those x 's for which $\|x_i\| \geq 1$ for some i together with the hyperplanes $\{x : |x_j| = \frac{2}{3} \text{ for some } j\}$. In particular, S_0 is closed.

Case 1. We denote by S_1 the set of x 's in $\mathbf{R}^n$ which belongs to at least one, but only to finitely many, *distinct* open cubes $Q_j(\sigma)$ corresponding to positive integers j and σ 's in Σ . (Recall that if $\sigma, \tilde{\sigma} \in \Sigma$ with $\sigma(j) = \tilde{\sigma}(j)$ for $j = 1, \dots, J$, then $Q_J(\sigma) = Q_J(\tilde{\sigma})$ so that these cubes are not distinct. Therefore, the word *distinct* in the specification of S_1 is very important.) Suppose then $x \in S_1$. In view of assertions (A) and (B) above, there must exist a unique positive integer K and unique vectors $\tau(1), \dots, \tau(K)$ in $\mathbf{R}^n$ (all depending on x) such that whenever $\sigma \in \Sigma$, $x \in \bigcap_{j=1}^K Q_j(\sigma)$ if and only if $\sigma(j) = \tau(j)$ for each $j = 1, \dots, K$. It follows, furthermore, that $x \notin Q_J(\sigma)$ whenever $J > K$ for any σ . Suppose then $\sigma \in \Sigma$ with $\sigma(j) = \tau(j)$ for each $j = 1, \dots, K$. We then define

$$f(x) = \sum_{j=1}^K 3^{-nj} F_j^{(\sigma)} \left(\frac{x - C_j(\sigma)}{L_j} \right);$$

here

$$F_j^{(\sigma)}(x) = \begin{cases} M_j^0(x) & \text{if } N_j(\sigma) = 0, \\ N_j(\sigma)M_j(x) & \text{if } N_j(\sigma) \geq 1. \end{cases}$$

This expression is well defined since it depends only on K and $\tau(1), \dots, \tau(K)$. Note that for such an x as just described, it is trivially true that

$$f(x) = \sum_{j=1}^J 3^{-nj} F_j^{(\sigma)}\left(\frac{x - C_j(\sigma)}{L_j}\right)$$

whenever $J > K$ and $\sigma(j) = \tau(j)$ for $j = 1, \dots, K$.

In Case ∞ below, we will show that the set S_∞ is compact. It will then follow that S_1 is open.

Case ∞ . We denote by S_∞ the set of x 's in $\mathbf{R}^n$ which belong to infinitely many distinct $Q_j(\sigma)$'s. In view of assertions (A), (B), and (C) above, we know for each positive integer j that

$$\bigcup_{\sigma \in \Sigma} Q_j(\sigma) \supset \bigcup_{\sigma \in \Sigma} \text{Clos}[P_j(\sigma)] \supset \bigcup_{\sigma \in \Sigma} P_j(\sigma) \supset \bigcup_{\sigma \in \Sigma} Q_{j+1}(\sigma)$$

and further infer that S_∞ has the Cantor set type representation

$$S_\infty = \bigcap_{j=1}^{\infty} \bigcup_{\sigma \in \Sigma} Q_j(\sigma) = \bigcap_{j=1}^{\infty} \bigcup_{\sigma \in \Sigma} P_j(\sigma) = \bigcap_{j=1}^{\infty} \bigcup_{\sigma \in \Sigma} \text{Clos}[P_j(\sigma)].$$

It follows that S_∞ is compact (and thus that S_1 is open as asserted above). Also as a consequence of our assertions, we know for each x in S_∞ that there is a unique σ in Σ such that $x \in \bigcap_{j=1}^{\infty} Q_j(\sigma)$; indeed, by assertion (B), $\{x\} = \bigcap_{j=1}^{\infty} Q_j(\sigma) = \bigcap_{j=1}^{\infty} P_j(\sigma)$. We define

$$f(x) = \sum_{j=1}^{\infty} 3^{-nj} F_j^{(\sigma)}\left(\frac{x - C_j(\sigma)}{L_j}\right) = \sum_{j=1}^{\infty} N_j(\sigma) 3^{-nj},$$

where the $F_j^{(\sigma)}$'s are defined as in Case 1. The final equality follows from the fact that when x belongs to $P_j(\sigma)$, as it must if $x \in S_\infty$, then $M_j(\dots) = 1$ for $N_j(\sigma) \neq 0$ while $M_j^0(\dots) = 0$ for $N_j(\sigma) = 0$.

Step 3. We shall now compute the derivatives of f up to order $n-1$. Case 0 is the easiest. If x is in the interior of S_0 , all derivatives are clearly zero because f vanishes identically on S_0 . The boundary of S_0 is also the boundary of $\bigcup_{\sigma \in \Sigma} Q_1(\sigma)$, and all derivatives of f are also zero there because all the derivatives of $M_1(x)$ are zero when $x \in \partial Q_0$.

The next easiest case is Case 1. Suppose then $x \in S_1$. If $x \in \bigcap_{j=1}^J Q_j(\sigma)$ for some σ , but does not belong to any $Q_{J+1}(\sigma)$, it can happen that x is in the boundary of $Q_{J+1}(\sigma)$, but x can never be in the boundary of $Q_{J+2}(\sigma)$. To investigate the behavior of $f(z)$ for z in a small neighborhood of x , it suffices to use the following formula which is valid in such a neighborhood:

$$f(z) = \sum_{j=1}^{J+1} 3^{-nj} F_j^{(\sigma)}\left(\frac{z - C_j(\sigma)}{L_j}\right).$$

From this it is clear that f is infinitely differentiable at x . If we differentiate f one or more times, then only the term with $j = J$ will contribute. The reason is that $F_k^{(\sigma)}$ is constant near x for $k < J$ (assertion (B)), and if x is on the boundary of $Q_{J+1}(\sigma)$, the derivatives of $F_{J+1}(\sigma)$ all vanish at x by continuity. Thus, for each positive integer k not exceeding $n - 1$, we have

$$|D^{(k)}f(x)| \leq (3^n - 1)3^{-nJ}A \left(\frac{1}{\delta\gamma^J L_J} \right)^k;$$

here $D^{(k)}f$ symbolically denotes any of the partial derivatives (in the coordinate directions) of f of order k . We have used property (e) above and our definitions of the M_j^0 's and M_j 's. We define

$$B = \left(\prod_{m=1}^{\infty} (1 - \delta\gamma^m) \right)^{-1} < \infty$$

and check that for each positive integer K ,

$$L_K = 3^{-K} \prod_{j=1}^{K-1} (1 - \delta\gamma^j) > 3^{-K} \prod_{j=1}^{\infty} (1 - \delta\gamma^j) = \frac{3^{-K}}{B}.$$

Note that B can be chosen to satisfy the inequality $B^{-n} \geq 1 - \alpha$ for any fixed choice of γ merely by choosing δ to be small enough; this is crucial for conclusion (5). Thus, for k 's not exceeding $n - 1$,

$$|D^{(k)}f(x)| \leq A\delta^{-k}B^k3^n3^{-nJ}(3/\gamma)^{Jk},$$

and this is bounded independent of J (and hence x) for all $k \leq n - 1$ provided $\gamma^{n-1} \geq \frac{1}{3}$, as we have assumed. As an important point, we note here that as a consequence of our construction, the set of points in S_1 on which $\nabla f = 0$ lies within a countable union of affine hypersurfaces in $\mathbf{R}^n$ and, in particular, has zero $\mathcal{L}^n$ measure. Furthermore, $\partial S_0 \cap \partial S_1 = \bigcup_{\sigma \in \Sigma} \partial Q_1(\sigma)$, and all the derivatives of f vanish on this common boundary. Thus, f is infinitely differentiable within the open set $S_1 \cup S_0$, and $D^{n-1}f$ is bounded on $S_1 \cup S_0$.

The next fact to be established is that all derivatives of f up to order $n - 1$ go to zero as $x \in S_1$ approaches S_{∞} , and that they do so at a rate which is some power of the L^{∞} distance from x to S_{∞} . Let $x \in S_1$ and $\sigma \in \Sigma$ such that $x \in Q_J(\sigma)$ but $x \notin Q_{J+1}(\tilde{\sigma})$ for any $\tilde{\sigma}$. Let $z \in S_{\infty}$ and $\tilde{\sigma} \in \Sigma$ so that $z \in P_{J+1}(\tilde{\sigma})$. Since $x \notin Q_{J+1}(\tilde{\sigma})$, the distance $\|x - z\|$ is bounded below by the distance of $P_{J+1}(\tilde{\sigma}) \subset Q_{J+1}(\tilde{\sigma})$ to the boundary of $Q_{J+1}(\tilde{\sigma})$. Thus,

$$\|x - z\| \geq \delta L_{J+1} \gamma^{J+1} \geq \delta(\gamma/3)^{J+1}/B.$$

We recall that $3 > 3\gamma^{n-1} > 1$ and define the number $1 > \lambda > 0$ by

$$\left(\frac{3}{\gamma}\right)^{\lambda} = 3\gamma^{n-1}$$

so that $(\frac{\gamma}{3})^{\lambda+n-1} = 3^{-n}$. This number λ , which will turn out to be the exponent of Hölder continuity, can be made as close to 1 as desired simply by choosing γ close to 1. Inserting this expression for 3^{-n} in our previous bound, we find that

$$\begin{aligned} |D^{(k)} f(x)| &\leq A\delta^{-k} B^k 3^n [(\gamma/3)^{\lambda+n-1-k}]^J \\ &\leq A\delta^{-k} B^k 3^n \left[\left(\frac{3B}{\gamma\delta} \right) \|x - z\| \right]^{\lambda+n-1-k}. \end{aligned}$$

We therefore conclude the existence of a finite constant C_1 such that whenever $x \in S_1$, $z \in S_\infty$, and k does not exceed $n-1$, then

$$|D^{(k)} f(x)| \leq C_1 \|x - z\|^{\lambda+n-1-k}.$$

Our last goals are to show that $D^{(k)} f(x) = 0$ when $x \in S_\infty$ and $k \leq n-1$ and also to show that $D^{(n-1)} f$ is Hölder continuous. In order to verify the first assertion for $k=1$, it suffices to show the existence of a function $\omega(t)$ for which $\omega(t)/t \rightarrow 0$ as $t \rightarrow 0$ such that

$$|f(x+z) - f(x)| \leq \omega(\|z\|)$$

whenever $x \in S_\infty$ and $\|z\| < \delta\gamma/3$. Fix a positive integer J so that $\delta\gamma^{J+1}L_{J+1} < \|z\| \leq \delta\gamma^J L_J$, and also fix $\sigma \in \Sigma$ so that $x \in P_J(\sigma)$. Thus,

$$\left\| x - 2 \sum_{j=1}^J C_j(\sigma) L_j \right\| < L_J(1 - \delta\gamma^J).$$

We set $\Gamma = \sum_{j=1}^J N_j(\sigma) 3^{-nj}$ and estimate $\Gamma \leq f(x) \leq \Gamma + 3^{-nJ}$. On the other hand,

$$\left\| x + z - 2 \sum_{j=1}^J C_j(\sigma) L_j \right\| < L_J(1 - \delta\gamma^J) + \delta\gamma^J L_J = L_J.$$

Hence, the point $x+z$ belongs to $Q_J(\sigma)$, and, therefore, whether $x+z$ belongs to S_1 or to S_∞ , we can estimate $\Gamma - 3^{-nJ}(3^n - 1) \leq f(x+z) \leq \Gamma + 3^{-nJ}$. Hence,

$$\begin{aligned} |f(x+z) - f(x)| &\leq 3^{-nJ} + 3^{-nJ}(3^n - 1) \\ &= 3^{-nJ+n} \\ &\leq 3^n \left(\frac{3B}{\delta\gamma} \right)^{\lambda+n-1} \|z\|^{\lambda+n-1}. \end{aligned}$$

The last inequality follows from our assumption that

$$\|z\| > \delta\gamma^{J+1}L_{J+1} \geq (\delta\gamma/3B)(\gamma/3)^J$$

and $(\gamma/3)^{\lambda+n-1} = 3^{-n}$. Thus, we have established our criterion with $\omega(t) = C_2 t^{\lambda+n-1}$.

For the higher derivatives, we proceed inductively. Assume that $1 \leq k \leq n-2$ and $D^{(j)}f(x) = 0$ whenever $1 \leq j \leq k$ and $x \in S_\infty$. Assume also (as shown above) that $|D^{(j)}f(x)| \leq C_1 \|x - z\|^{\lambda+n-1-j}$ whenever $1 \leq j \leq n-1$ and x belongs to S_1 and z to S_∞ . We must then show that whenever x belongs to S_∞ and z is arbitrary (but small), then

$$|D^{(k)}f(x) - D^{(k)}f(x+z)| = |D^{(k)}f(x+z)| \leq \omega(\|z\|)$$

with $\omega(t)/t \rightarrow 0$ as $t \rightarrow 0$. If $x+z \in S_\infty$, the inequality is trivial. If $x+z \in S_1$,

$$|D^{(k)}f(x+z)| \leq C_1 \|(x+z) - x\|^{\lambda+n-1-k}$$

which satisfies our desideratum.

The same inequality implies that $D^{(n-1)}f$ is Hölder continuous with exponent λ . We have shown that

$$|D^{(n-1)}f(x) - D^{(n-1)}f(z)| \begin{cases} = 0 & \text{if } x \in S_\infty \text{ and } z \in S_\infty, \\ \leq C_1 \|x - z\|^\lambda & \text{if } x \in S_\infty \text{ and } z \in S_1 \end{cases}$$

while $f \in C^\infty(S_1 \cup S_0)$ and $D^{(n-1)}f$ is bounded on $S_1 \cup S_0$. The Hölder continuity then follows from the compactness of Q_0 and the fact that $D^{(n-1)}f(x) = 0$ for $x \notin Q_0$.

Step 4. We have shown that apart from a set of zero $\mathcal{L}^n$ measure, the set on which $\nabla f = 0$ and $f > 0$ is precisely S_∞ . Therefore,

$$\mathcal{G}_f(y) = \mathcal{L}^n(S_\infty \cap \{x : f(x) > y\}).$$

Let us first compute $\mathcal{L}^n(S_\infty) = \mathcal{G}_f(0)$. To do this, first define for positive integers J ,

$$V_J = \bigcup_{\sigma \in \Sigma} P_J(\sigma).$$

There are exactly 3^{Jn} inner cubes $P_J(\sigma)$ in V_J , and, therefore,

$$\mathcal{L}^n(V_J) = 3^{Jn} [2L_J(1 - \delta\gamma^J)]^n = 2^n \prod_{j=1}^J (1 - \delta\gamma^j)^n$$

since $L_J = 3^{-j} \prod_{j=1}^{J-1} (1 - \delta\gamma^j)$. Clearly, $V_J \subset V_{J+1}$ for each J . Hence,

$$\begin{aligned} \mathcal{G}_f(0) &= \mathcal{L}^n(S_\infty) \\ &= \lim_{J \rightarrow \infty} 2^n \prod_{j=1}^J (1 - \delta\gamma^j)^n \\ &= 2^n \prod_{j=1}^{\infty} (1 - \delta\gamma^j)^n \\ &= (2/B)^n > 0. \end{aligned}$$

As noted before, by choosing δ small enough,

$$\mathcal{G}_f(0) = \mathcal{L}^n(Q_0 \cap \{x : \nabla f(x) = 0\}) > (1 - \alpha)2^n,$$

which is conclusion (5).

The final thing to check is that whenever $0 \leq y < z \leq 1$, then

$$\mathcal{L}^n(S_\infty \cap \{x : y < f(x) \leq z\}) = \mathcal{G}_f(y) - \mathcal{G}_f(z) = C(z - y)$$

with $C = \mathcal{L}^n(S_\infty)$. Equivalently (see Theorem 3.2), we wish to show that the measure $f_\#(\mathcal{L}^n \llcorner S_\infty)$ equals $\mathcal{L}^n(S_\infty) \cdot \mathcal{L}^1 \llcorner [0, 1]$; here, for each Borel subset A of $\mathbf{R}$, $f_\#(\mathcal{L}^n \llcorner S_\infty)(A) = \mathcal{L}^n(S_\infty \cap f^{-1}(A))$. To see this, we fix $\sigma_* \in \Sigma$, and for each positive integer J , set

$$\Sigma_J = \Sigma \cap \{\sigma : \sigma(j) = \sigma_*(j) \text{ for each } j > J\}$$

so that $\text{card}(\Sigma_J) = 3^{nJ}$. Our Σ_J is an index set for J th generation cubes. For $p \in [0, 1]$, we denote by $\llbracket p \rrbracket$ a unit point mass measure located at p , and for each positive integer J , set

$$\mu_J = \sum_{\sigma \in \Sigma_J} \mathcal{L}^n(P_J(\sigma)) \cdot \left[\sum_{j=1}^J N_j(\sigma) 3^{-nJ} \right].$$

Each μ_J thus consists of equal point masses uniformly distributed along the unit interval. It readily follows from the construction of our f and the estimates above that $\mu_J \rightarrow f_\#(\mathcal{L}^n \llcorner S_\infty)$ weakly as measures as $J \rightarrow \infty$. This finishes our demonstration.

The five facts about f mentioned in the beginning have been verified. ■

5.2. Theorem (Sufficient differentiability implies co-area regularity). *Suppose $f : \mathbf{R}^n \rightarrow \mathbf{R}^+$ belongs to our class $\mathcal{G}_1$. If $f \in C_{\text{loc}}^{n-1,1}(\mathbf{R}^n)$, i.e., f is $(n-1)$ times continuous differentiable and its $(n-1)$ st derivatives are locally Lipschitz, then f is co-area regular.*

Proof. First, assume $f \in C^n(\mathbf{R}^n)$ and apply the general Morse-Sard-Federer theorem as stated in [FH, 3.4.3] with m replaced by n , k replaced by n , A replaced by $\mathbf{R}^n$, ν replaced by 0, Y replaced by $\mathbf{R}$, and $B = \{x : \nabla f(x) = 0\}$. We conclude that $\mathcal{L}^1(f(B)) = 0$, since Hausdorff's measure $\mathcal{H}^1$ equals Lebesgue's measure $\mathcal{L}^1$ on $\mathbf{R}$.

By Definition 1.2.5 and Theorem 3.2, for any set $U \subset \mathbf{R}^+$,

$$\mathcal{G}'_f(U) = - \int \chi_{\{f \in U\}} \chi_{\{\nabla f = 0\}} d\mathcal{L}^n \geq - \int \chi_{\{f \in f(B)\}} \chi_{\{\nabla f = 0\}} d\mathcal{L}^n.$$

But $\mathcal{L}^1(f(B)) = 0$, and so the measure $\mathcal{G}'_f$ is carried on a set of zero Lebesgue measure and is thus singular.

In case $f \in C_{\text{loc}}^{n-1,1}(\mathbf{R}^n)$, we can apply the theorem [FH, 3.1.15] with m replaced by n , n replaced by 1, k replaced by $n-1$, and with $A = B$ being bounded open sets in $\mathbf{R}^n$. We conclude that f can be changed on a set of arbitrarily small measure to become $C^n(\mathbf{R}^n)$ on $\mathbf{R}^n$. It follows readily from the $C^n(\mathbf{R}^n)$ case above that $\mathcal{G}'_f$ has no absolutely continuous part. ■

5.3. Theorem (Co-area regular and co-area irregular functions are dense in $\mathcal{E}_1$). Suppose $n \geq 2$ and $\mathcal{E}$ is a topological space of functions $f: \mathbf{R}^n \rightarrow \mathbf{R}^+$. Suppose the following:

- (a) $\mathcal{E}$ is contained in our class $\mathcal{E}_1$;
- (b) $C_c^1(\mathbf{R}^n)^+ \subset \mathcal{E}$ and $C_c^\infty(\mathbf{R}^n)^+$ is dense in $\mathcal{E}$;
- (c) if $g \in C_c^\infty(\mathbf{R}^n)^+$ and $h \in C_c^1(\mathbf{R}^n)^+$ with disjoint supports, then $g + h \in \mathcal{E}$ and $g + th \rightarrow g$ as $t \downarrow 0$.

Then each of the following two sets is dense in $\mathcal{E}$:

$$\mathcal{E} \cap \{f: f \text{ is co-area regular}\} \quad \text{and} \quad \mathcal{E} \cap \{f: f \text{ is co-area irregular}\}.$$

Proof. According to Theorem 5.2, each member of $C_c^\infty(\mathbf{R}^n)$ is co-area regular. Hypothesis (b) then implies that $\mathcal{E} \cap \{f: f \text{ is co-area regular}\}$ is dense in $\mathcal{E}^+$.

We will now show that each member of $C_c^\infty(\mathbf{R}^n)^+$ can be approximated by a member of $\mathcal{E} \cap \{f: f \text{ is co-area irregular}\}$. This will imply $\mathcal{E} \cap \{f: f \text{ is co-area irregular}\}$ is also dense in $\mathcal{E}$. Suppose that $g \in C_c^\infty(\mathbf{R}^n)^+$ and that $f \in C_c^1(\mathbf{R}^n)^+$ is one of the nonnegative co-area irregular functions constructed in 5.1. Let h be a suitable translate of f so that h and g have disjoint supports. For each $t > 0$, the function $g + th$ is co-area irregular because $\mathcal{V}_{g+th} = \mathcal{V}_g + \mathcal{V}_{th}$. Hypothesis (c) implies that $g + th$ converges to g as $t \downarrow 0$. ■

Finally, we state and prove the theorem that shows there is no conflict between our results in this paper and the earlier result of Coron [CJ] about the continuity of rearrangement in one dimension.

5.4. Theorem (If $n = 1$, there are no co-area irregular functions). If our domain dimension n equals 1, then each function in $\mathcal{E}_1$ (see Definition 1.2.3) is co-area regular (see Definition 1.2.6).

Proof. For a function $f: \mathbf{R} \rightarrow \mathbf{R}^+$ which belongs to $\mathcal{E}_1$ and for $\sigma > 0$, let $f_\sigma = \max\{0, f - \sigma\}$. It is clear that if f is not co-area regular, then some f_σ is also not co-area regular. Hence it suffices to show that $f \in \mathcal{E}_1$ is co-area regular under the additional assumption that f and f' vanish in the complement of a set Ω of finite $\mathcal{L}^1$ measure and that f' is $\mathcal{L}^1$ summable. Fix such a function f .

For each $\varepsilon > 0$, we will show the existence of a set B in $\mathbf{R}^+$ which is the union of countably many closed intervals together with $\{0\}$ such that $\mathcal{L}^1(B) < \varepsilon$ and

$$\mathcal{L}^1(\{x: f'(x) = 0\} \sim f^{-1}(B)) = 0.$$

This implies, by letting ε tend to zero, that $f(\{x: f(x) > 0 \text{ and } f'(x) = 0\})$ has zero Lebesgue measure. This clearly implies the co-area regularity of f . Now fix $\varepsilon > 0$.

Recall that a point a in $\mathbf{R}$ belongs to the Lebesgue set of the summable function f' provided

$$\lim_{r \downarrow 0} \frac{1}{2r} \int_{x \in [a-r, a+r]} |f'(x) - f'(a)| d\mathcal{L}^1 x = 0$$

and that $\mathcal{L}^1$ almost every point of $\mathbf{R}$ belongs to the Lebesgue set of f' [FH, 2.9.9]. We let A denote the collection of those points in Ω which belong to the Lebesgue set of f' and at which f' vanishes. Clearly, $\mathcal{L}^1(A) \leq \mathcal{L}^1(\Omega) < \infty$. If $\mathcal{L}^1(A) = 0$, there is nothing to prove because then $f'(x) \neq 0$ for almost every x in Ω . Suppose then $\mathcal{L}^1(A) > 0$ and, for each a in A , let $0 < \rho(a) < 1$ be chosen such that

$$(1) \quad \frac{1}{2r} \int_{x \in [a-r, a+r]} |f'(x) - f'(a)| d\mathcal{L}^1 x = \frac{1}{2r} \int_{x \in [a-r, a+r]} |f'(x)| d\mathcal{L}^1 x < \frac{\varepsilon}{2\mathcal{L}^1(A)}$$

whenever $0 < r < \rho(a)$. Now use Besicovitch's covering theorem [FH, 2.8.15] to choose $a_1, a_2, a_3, \dots \in A$ and numbers $r_1, r_2, r_3, \dots$ with $0 < r_i < \rho(a_i)$ for each i such that $|a_j - a_k| > r_j + r_k$ whenever $j \neq k$ and

$$\mathcal{L}^1(A \setminus \bigcup_{i=1}^{\infty} [a_i - r_i, a_i + r_i]) = 0.$$

In particular, $\mathcal{L}^1(A) = \sum_{i=1}^{\infty} 2r_i < \infty$. We set $b_i = f(a_i)$ and

$$(2) \quad s_i = \int_{[a_i - r_i, a_i + r_i]} |f'| d\mathcal{L}^1$$

for each i and infer from (2) and the fundamental theorem of calculus (in the sense of distributions) that $f([a_i - r_i, a_i + r_i]) \subset [b_i - s_i, b_i + s_i]$. Using (1) and (2) together, we conclude

$$(3) \quad \mathcal{L}^1[b_i - s_i, b_i + s_i] = 2s_i \leq \frac{\varepsilon}{2\mathcal{L}^1(A)} 2r_i.$$

We set $B = \{0\} \cup \bigcup_{i=1}^{\infty} [b_i - s_i, b_i + s_i]$, and it follows trivially from the covering construction above that $f'(x) \neq 0$ for $\mathcal{L}^1$ almost every x which does not belong to B . By (3), we have

$$\mathcal{L}^1(B) \leq \sum_{i=1}^{\infty} 2s_i \leq \frac{\varepsilon}{\mathcal{L}^1(A)} \sum_{i=1}^{\infty} 2r_i = \varepsilon,$$

and thus our B satisfies the requirements mentioned at the beginning of the proof. ■

§6. REARRANGEMENT IS NOT CONTINUOUS AT CO-AREA IRREGULAR FUNCTIONS

Theorem 4.3 states that when $f \in \mathcal{E}_1$ is co-area irregular, one can always find a sequence of functions f_j in $\mathcal{E}_1$ such that the ∇f_j 's converge to ∇f in several strong senses (which includes all interesting topologies) while the ∇f_j^* 's do not converge to ∇f^* in $L^1_{\text{loc}}(\mathbf{R}^n, \mathbf{R}^n)$. The proof will now be given.

6.1. Proof of Theorem 4.3. Step 1. Suppose $0 < Y < Z < \infty$ as in Theorem 4.3. The essence of co-area irregularity is the presence of an absolutely continuous part of the derivative of the residual distribution function $\mathcal{G}_f$ associated with the set $A^{(0)} = \{x : \nabla f(x) = 0 \text{ and } Y < f(x) < Z\}$. We show in this step how $A^{(0)}$ can be reduced to a smaller, compact set $A^{(1)}$ which also contributes an absolutely continuous part to this derivative. By analogy with the definition of $\mathcal{G}_f$, we associate with any $\mathcal{L}^n$ measurable subset B of $A^{(0)}$ the function $\mathcal{G}_B : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ specified by setting

$$\mathcal{G}_B(y) = \int \chi_{\{f > y\}} \chi_B d\mathcal{L}^n.$$

Clearly, $\mathcal{G}_f$, $\mathcal{G}_B$, and $[\mathcal{G}_f - \mathcal{G}_B]$ are monotonically nonincreasing. Hence, there are nonnegative $\mathcal{L}^1$ summable functions h , h_B , and $\tilde{h}$ with $h = h_B + \tilde{h}$ and Radon measures ν , ν_B , and $\tilde{\nu}$ with $\nu = \nu_B + \tilde{\nu}$ which are singular with respect to $\mathcal{L}^1$ such that, in the sense of distributions, the derivatives of $-\mathcal{G}_f$, $-\mathcal{G}_B$, and $-(\mathcal{G}_f - \mathcal{G}_B)$ are respectively $\mathcal{L}^1 \wedge h + \nu$, $\mathcal{L}^1 \wedge h_B + \nu_B$, and $\mathcal{L}^1 \wedge \tilde{h} + \tilde{\nu}$. The same reasoning implies that if C is another measurable subset of $A^{(0)}$ with $B \subset C$, then $h_B \leq h_C$ and $\nu_B \leq \nu_C$. Since ν is singular with respect to $\mathcal{L}^1$, there is a subset E of $\mathbf{R}^+$ of full $\mathcal{L}^1$ measure such that $\nu(E) = 0$ (hence, also, $\nu_B = 0$ for any B). It follows that

$$G \equiv \int \chi_{\{f \in (Y, Z) \cap E\}} \chi_{\{\nabla f = 0\}} d\mathcal{L}^n = \int_{(Y, Z)} h d\mathcal{L}^1.$$

Suppose that $B_1 \subset B_2 \subset B_3 \cdots$ are measurable subsets of $A^{(0)}$ such that

$$\mathcal{L}^n(A^{(0)} \setminus B_j) \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

Then for the same reason,

$$G_j \equiv \int \chi_{\{f \in (Y, Z) \cap E\}} \chi_{B_j} \chi_{\{\nabla f = 0\}} d\mathcal{L}^n = \int_{(Y, Z)} h_{B_j} d\mathcal{L}^1.$$

Both sides of this equation increase with j and $\lim_{j \rightarrow \infty} G_j = G$ by the increasing convergence theorem for measurable sets. Since h_{B_j} is monotone nondecreasing in j , it has a pointwise limit $\hat{h} \leq h$. But by monotone convergence

$$\begin{aligned} \int_{(Y, Z)} h d\mathcal{L}^1 &= G = \lim_{j \rightarrow \infty} G_j \\ &= \lim_{j \rightarrow \infty} \int_{(Y, Z)} h_{B_j} d\mathcal{L}^1 \\ &= \int_{(Y, Z)} \hat{h} d\mathcal{L}^1 \leq \int_{(Y, Z)} h d\mathcal{L}^1. \end{aligned}$$

Therefore, $\lim_{j \rightarrow \infty} h_{B_j}(y) = h(y)$ for almost every $y \in (Y, Z)$.

As an application of this remark, we take B_j to be $A^{(0)} \cap U_j$ where U_j is the ball of radius j centered at $0 \in \mathbf{R}^n$. For some J , we have that

$$\int_{(3Y+Z)/4}^{(3Z+Y)/4} h_{B_j} d\mathcal{L}^1 > 0;$$

we fix such a J . Since $A^{(0)} \cap U_j$ is $\mathcal{L}^n$ measurable with finite measure, it contains a compact subset $A^{(1)}$ for which $\int_{(3Y+Z)/4}^{(3Z+Y)/4} h_{A^{(1)}} d\mathcal{L}^1 > 0$; we further set $h^{(1)} = h_{A^{(1)}}$. At the end of this proof, we shall have occasion to decrease $A^{(0)}$ still further to a smaller set, but for now we remain with $A^{(1)}$ which is compact. This property will be needed to prove property (3) of the sequence f_j which we shall now construct.

Step 2. Fix some rotationally invariant function $K: \mathbf{R}^n \rightarrow \mathbf{R}^+$ which is infinitely differentiable and has compact support and for which $\int K d\mathcal{L}^n = 1$. For each positive integer j , then set $K_j(x) = j^{n/2} K(j^{1/2}x)$. Let $\chi^{(1)}$ be the characteristic function of the set $A^{(1)}$, and consider the convolutions $W_j = K_j * \chi^{(1)}$. This sequence of functions $W_j: \mathbf{R}^n \rightarrow \mathbf{R}^+$ has the following properties.

- (a) Each W_j is infinitely differentiable with $0 \leq W_j(x) \leq 1$ for each x .
- (b) For each j , both W_j and ∇W_j are $\mathcal{L}^n$ summable with

$$\|W_j\|_{L^1(\mathbf{R}^n)} = \|\chi^{(1)}\|_{L^1(\mathbf{R}^n)} \leq \mathcal{L}^n\{x: f(x) > Y\}$$

and

$$\|\nabla W_j\|_{L^1(\mathbf{R}^n, \mathbf{R}^n)} \leq j^{1/2} C_1 \quad \text{with } C_1 = \|\chi^{(1)}\|_{L^1(\mathbf{R}^n)} \|\nabla K\|_{L^1(\mathbf{R}^n)}.$$

- (c) For each x , we have the pointwise estimate $|\nabla W_j(x)| \leq C_2 j^{1/2}$ with $C_2 = \|\nabla K\|_{L^1}$.

(d) The W_j 's converge to $\chi^{(1)}$ in $L^1(\mathbf{R}^n)$, and $W_j(x) \rightarrow \chi^{(1)}(x)$ for almost every x .

(e) Let R denote the radius of the support of K . Then the support of W_j lies within the set $\{x+y: x \in A^{(1)} \text{ and } |y| \leq j^{-1/2}R\}$. From this we conclude, using the fact that $A^{(1)}$ is compact, that $\mathcal{L}^n\{x: W_j(x) > 0 \text{ and } x \notin A^{(1)}\}$ converges to zero as $j \rightarrow \infty$.

Step 3. Let $F: \mathbf{R} \rightarrow [0, 1]$ be an infinitely differentiable function such that $F(t) = 1$ for $(4Y+Z)/5 \leq t \leq (4Z+Y)/5$ and $F(t) = 0$ for $t \leq (5Y+Z)/6$ or $t \geq (5Z+Y)/6$. Define $S_j: \mathbf{R} \rightarrow \mathbf{R}$ by setting for each t

$$S_j(t) = j^{-1} F(t) \sin(tj),$$

and let $1/C_3 = 2 \sup_{t,j} |\dot{S}_j(t)|$. (The dot denotes derivative.) We define our required sequence of functions by setting

$$f_j(x) = f(x) + C_3 W_j(x) S_j(f(x))$$

for each $j = 1, 2, 3, \dots$. Since the function

$$g_w(t) \equiv t + C_3 w S_j(t)$$

is strictly monotone increasing in t when $0 \leq w \leq 1$, we conclude that $f_j(x) \geq 0$ for each x and each j . Clearly, all the f_j 's belong to $\mathbf{L}_{\text{loc}}^1(\mathbf{R}^n)$, and they converge to f in $\mathbf{L}^\infty(\mathbf{R}^n)$. Also, it is clear that $f_j(x) = f(x)$ if $f(x) \notin (Y, Z)$, and, in view of our definition of F , that $f_j(x) \in (Y, Z)$ if $f(x) \in (Y, Z)$ and j is large enough. In particular, the conditions specified in (1) of Theorem 4.3 for the f_j 's are satisfied.

In order to establish the conditions specified in (2) of Theorem 4.3 about gradient behavior, we compute

$$\begin{aligned} \nabla f_j(x) - \nabla f(x) &= C_3 W_j(x) F(f(x)) \cos(jf(x)) \nabla f(x) \\ &\quad + j^{-1} C_3 W_j(x) \dot{F}(f(x)) \sin(jf(x)) \nabla f(x) \\ &\quad + C_3 S_j(f(x)) \nabla W_j(x). \end{aligned}$$

We wish to show that $\nabla f_j - \nabla f$ converges to zero in $\mathbf{L}^1(\mathbf{R}^n, \mathbf{R}^n)$ norm. The second term on the right side goes to zero in $\mathbf{L}^1(\mathbf{R}^n, \mathbf{R}^n)$ norm according to the dominated convergence theorem since $C_3 W_j \dot{F} \sin$ is bounded and since $j^{-1} \rightarrow 0$ as $j \rightarrow \infty$. The third term also goes to zero by the dominated convergence theorem since $S_j \leq j^{-1}$ and since $|\nabla W_j| \leq j^{1/2} C_2$. Since the first term is clearly dominated by a summable function, it suffices to show its pointwise convergence to zero almost everywhere. According to property (d) above, $W_j \rightarrow \chi^{(1)}$ almost everywhere as $j \rightarrow \infty$ and $\chi^{(1)} \nabla f = 0$ by construction. It is now immediate that the conditions specified in (2) of Theorem 4.3 are satisfied for all large j . Finally, condition (3) of Theorem 4.3 is implied by (e) above.

Step 4. In this step, we assume that j is fixed. To study the rearrangements f_j^* and f^* , we consider the distribution functions

$$\mathcal{V}_j(y) \equiv \mathcal{V}_{f_j}(y) = \int \chi_{\{f_j > y\}} d\mathcal{L}^n$$

and

$$\mathcal{V}(y) \equiv \mathcal{V}_f(y) = \int \chi_{\{f > y\}} d\mathcal{L}^n.$$

Suppose m is a fixed integer, $y = 2\pi m/j$, and $x \in \mathbf{R}^n$. We assert

- (i) if $f(x) = y$, then $f_j(x) = y$;
- (ii) if $f(x) > y$, then $f_j(x) > y$;
- (iii) if $f(x) < y$, then $f_j(x) < y$.

Assertion (i) is clear since $\sin(2\pi m) = 0$. To check assertion (ii), we suppose $f(x) = y + \varepsilon$ with $\varepsilon > 0$, and set $w = W_j(x)$. We use the definition of f_j and the definition and monotonicity of g_w to infer

$$f_j(x) = g_w(y + \varepsilon) > g_w(y) = y$$

as required. A similar estimate shows assertion (iii). As a consequence of (i), (ii), and (iii), we conclude $\{x : f(x) > y\} = \{x : f_j(x) > y\}$ so that $\mathcal{V}(y) = \mathcal{V}_j(y)$.

Next consider the case in which m is an integer, $0 < \sigma < 1$, and $y = j^{-1}\pi(2m + \sigma)$. Assume also

$$(4Y + Z)/5 \leq j^{-1}\pi(2m - 1) \leq j^{-1}\pi(2m + 1) \leq (4Z + Y)/5,$$

in which case $F(y - t) = 1$ for $0 \leq t \leq \pi/j$. This latter condition will be automatically satisfied in $j \geq 20\pi(Z - Y)^{-1}$ (which will henceforth be assumed) and if $(3Y + Z)/4 \leq y \leq (3Z + Y)/4$.

We claim that $f_j(x) > y$ whenever the two conditions $0 < \sigma < 1$ and

$$f(x) > y - 2j^{-1}C_4W_j(x)\sin(\pi\sigma) \quad \text{with } 2C_4 = (1 + 1/C_3 + 2/\pi)^{-1} < 1$$

are satisfied. The truth of this claim implies that when $(3Y + Z)/4 \leq y \leq (3Z + Y)/4$, then

$$\mathcal{V}_j(y) \geq \mathcal{V}(y) + \delta_j(y)$$

with

$$\delta_j(y) = \int \chi_{\{y \geq f > y - j^{-1}C_4\sin(\pi\sigma)\}} \chi_{\{W_j > 1/2\}} d\mathcal{L}^n.$$

If $W_j(x) = 0$, our claim is immediate. Henceforth, we assume $W_j(x) > 0$. Also if $f(x) = y$, then clearly $f_j(x) \geq y$ since $\sin(yj) > 0$. On the other hand, if $f(x) > y$, the strict monotonicity of $g_w(t)$ with $w = W_j(x)$ again readily implies $f_j(x) > y$. For the remaining case, we suppose $f(x) = y - \gamma/j$ with $0 \leq \gamma < 2C_4W_j(x)\sin(\pi\sigma)$. Since $2C_4 < 1 < \frac{\pi}{2}$, we have that $F(y) = F(y - \gamma/j) = 1$. We have to verify that

$$f_j(x) - y = -j^{-1}\gamma + j^{-1}C_3W_j(x)\sin(\sigma\pi - \gamma) > 0.$$

Again, by the strict monotonicity of g_w , it suffices to consider the upper limiting value of γ and then to verify that

$$\begin{aligned} 2C_4\sin(\pi\sigma) &\leq C_3\sin[\pi\sigma - 2C_4W_j(x)\sin(\pi\sigma)] \\ &= C_3\sin(\pi\sigma)\cos(2C_4W_j(x)\sin(\pi\sigma)) \\ &\quad - C_3\cos(\pi\sigma)\sin(2C_4W_j(x)\sin(\pi\sigma)). \end{aligned}$$

Using the inequality $\cos t \geq 1 - \frac{2t}{\pi}$ for $0 \leq t \leq \frac{\pi}{2}$, we can bound the first term on the right from below (since $2C_4 < 1 < \frac{\pi}{2}$) by the quantity $C_3\sin(\pi\sigma) \times [1 - 4C_4/\pi]$. The second term can be bounded below by $-C_3[2C_4\sin(\pi\sigma)]$ since $\sin x \leq x$ for $x \geq 0$. The desired inequality is thus implied by the condition

$$2C_4 \leq C_3[1 - 4C_4/\pi] - C_3[2C_4]$$

which follows from the definition of C_4 above.

Step 5. We are finally ready to prove that $\nabla f_j^* \rightharpoonup \nabla f^*$ in $L^1(\mathbf{R}^n, \mathbf{R}^n)$. Let us abuse notation by writing $f_j^*(r)$ and $f^*(r)$ with $r = |x|$. We now choose and fix a sufficiently large integer J and two other large integers M_0

and $M_0 + N$ so that when we set $Y^* = 2\pi M_0/J$ and $Z^* \equiv 2\pi(M_0 + N)/J$, then $Y < Y^* < Z^* < Z$ and

$$I \equiv \int_{Y^*}^{Z^*} h^{(1)}(y) d\mathcal{L}^1 y > 0.$$

We also require that J be sufficiently large so that our various earlier conditions are all realized. Instead of using the sequence f_j with $j = J, J+1$, etc., we shall use the dyadic sequence $J, 2J, \dots, 2^k J, \dots$ and label this sequence $f_{(k)}$. Thus, for each k , our endpoints Y^* and Z^* are of the form $2\pi(\text{integer})/2^k J$, and there are $2^k N$ subintervals each of length $2\pi/2^k J$ lying in the interval between Y^* and Z^* .

For y 's in the range of f^* , we define $R(y)$ by requiring

$$\alpha(n)R(y)^n = \mathcal{V}(y),$$

and similarly define $R_{(k)}$'s associated with $f_{(k)}^*$'s. Then $f^*(r) = y$ whenever $r = R(y)$. We set $R_1 = R(Y^*)$ and $R_2 = R(Z^*)$. Clearly, $R_2 < R_1$ since $I > 0$, and we can also assume that $R_2 > 0$ (for otherwise take a slightly smaller value for Z^*). For each k , there are $N(k)+1$ special values of y (with $N(k) = 2^k N$) in the interval $[Y^*, Z^*]$. These values are $y_l^k = Y^* + 2\pi l/2^k J$, with $l = 0, 1, \dots, N(k)$. Let r_l^k be the corresponding radii, i.e., $\alpha(n)(r_l^k)^n = \mathcal{V}(y_l^k)$. Since $\mathcal{V}_k(y) = \mathcal{V}(y)$ at these special values, we have $f_{(k)}^*(r_l^k) = f^*(r_l^k)$ for each k and l .

We set $\partial f^* \equiv df^*(r)/dr$ and estimate

$$\begin{aligned} \int |\nabla f_{(k)}^* - \nabla f^*| d\mathcal{L}^n &\geq n\alpha(n) \int_{R_2}^{R_1} |\partial f_{(k)}^*(r) - \partial f^*(r)| r^{n-1} d\mathcal{L}^1 r \\ &\geq n\alpha(n) R_2^{n-1} \int_{R_2}^{R_1} |\partial f_{(k)}^*(r) - \partial f^*(r)| d\mathcal{L}^1 r \\ &\equiv n\alpha(n) R_2^{n-1} T_k. \end{aligned}$$

Our goal therefore is to bound T_k from below. For this purpose, consider the graphs of the functions f^* and $f_{(k)}^*$. (Recall that since ∂f^* and $\partial f_{(k)}^*$ are in $L_{\text{loc}}^1((R_2, \infty))$, these are continuous functions.) These graphs intersect at the special points (r_l^k, y_l^k) . Consider also the unsigned, two-dimensional area A between these two graphs for $R_2 \leq r \leq R_1$. Since $f_{(k)}^*(r)$ and $f^*(r)$ are monotone in r , we can use Fubini's theorem to compute A as two different integrals

$$A = \int_{Y^*}^{Z^*} |R_k(y) - R(y)| d\mathcal{L}^1 y = \int_{R_2}^{R_1} |f_{(k)}^*(r) - f^*(r)| d\mathcal{L}^1 r.$$

For $l = 0, 1, \dots, N(k)$, we define the subareas

$$A_k(l) = \int_{y_l^k}^{y_{l+1}^k} |R_k(y) - R(y)| d\mathcal{L}^1 y = \int_{r_{l+1}^k}^{r_l^k} |f_{(k)}^*(r) - f^*(r)| d\mathcal{L}^1 r.$$

These last two integrals are equal since the graphs intersect at the special values of y . We set $\mu_l^k = (r_l^k - r_{l+1}^k)^{-1}$ and conclude that

$$\begin{aligned} T_k(l) &\equiv \int_{r_{l+1}^k}^{r_l^k} |\partial f_{(k)}^*(r) - \partial f(r)| d\mathcal{L}^1 r \\ &\geq \mu_l^k \int_{r_{l+1}^k}^{r_l^k} |f_{(k)}^*(r) - f^*(r)| d\mathcal{L}^1 r \end{aligned}$$

since $f_{(k)}^*(r) - f^*(r)$ vanishes at the end points r_l^k and r_{l+1}^k .

Now $T_k = \sum_l T_k(l)$ and therefore

$$T_k \geq \sum_{l=0}^{N(k)} \mu_l^k A_k(l).$$

By the Schwarz inequality for sums and the fact that $\sum_l (\mu_l^k)^{-1} = R_1 - R_2$, we have that

$$T_k \geq (R_1 - R_2)^{-1} \left(\sum_{l=0}^{N(k)} A_k(l)^{1/2} \right)^2.$$

By the Schwarz inequality for integrals,

$$\left(\int_{y_l^k}^{y_{l+1}^k} d\mathcal{L}^1 y \right)^{1/2} (A_l^k)^{1/2} \geq \int_{y_l^k}^{y_{l+1}^k} |R_k(y) - R(y)|^{1/2} d\mathcal{L}^1 y.$$

Since $y_{l+1}^k - y_l^k = 2\pi/s^k J$ and does not depend on l , we infer

$$T_k \geq (R_1 - R_2)^{-1} (2^k J / 2\pi) \left(\int_{R_2}^{R_1} |R_k(y) - R(y)|^{1/2} d\mathcal{L}^1 y \right)^2.$$

When y lies in one of the $N(k)$ intervals $a_l^k = (2\pi l / 2^k J, (2\pi l + \pi) / 2^k J)$, we have proved a lower bound on $R_k(y) - R(y)$, namely,

$$\alpha(n)[(R_k(y))^n - (R(y))^n] \geq \delta_k(y)$$

with $y = \pi(2l + \sigma) / 2^k J$ and with $j = 2^k J$ in that formula. Since $R_k(y)$ and $R(y)$ do not exceed R_1 when $Y^* < y < Z^*$, we have that

$$R_k(y) - R(y) \geq [n\alpha(n)(R_1)^{n-1}]^{-1} \delta_k(y)$$

for y in one of the a_l^k intervals. To simplify matters, let us consider only the smaller intervals

$$b_l^k = [(2\pi l + \frac{\pi}{4}) / 2^k J, (2\pi l + \frac{3\pi}{4}) / 2^k J]$$

in which $\sin(\pi\sigma) \geq 2^{-1/2}$. We define their union $I^k = \bigcup_{l=0}^{N(k)} b_l^k \in (Y^*, Z^*)$, and we denote the characteristic function of I^k by χ_k . We then have for $y \in I^k$ that

$$\delta_k(y) \geq \nu_k(y) \equiv \int \chi_{\{y \geq f > y - j^{-1} C_3\}} \chi_{\{w_j > 1/2\}} d\mathcal{L}^n$$

with $C_5 = 2^{-1/2}C_4$ and $j = 2^k/J$. Using the above inequality for T_k , it suffices for us to prove that with

$$\Delta_k \equiv 2^{k/2} \int_{R_1}^{R_2} \chi_k(y) \nu_k(y)^{1/2} d\mathcal{L}^1 y,$$

we have

$$\limsup_{k \rightarrow \infty} \Delta_k > 0.$$

Recall that as $j \rightarrow \infty$, $W_j(x) \rightarrow \chi^{(1)}(x)$ for $\mathcal{L}^n$ almost every x , and therefore $\chi_{\{W_j > 1/2\}}(x) \rightarrow \chi^{(1)}(x)$ also for $\mathcal{L}^n$ almost every x . By Egoroff's theorem, there is a decreasing sequence of measurable sets $D_k \subset \mathbb{R}^n$ such that $\mathcal{L}^n(D_k) \rightarrow 0$ as $k \rightarrow \infty$ and such that

$$\lim_{j \rightarrow \infty} \chi^{(1)}(x) [\chi^{(1)}(x) - \chi_{\{W_j > 1/2\}}(x)] = 0 \quad \text{uniformly for all } x \notin D_k.$$

Returning to Step 1, let us consider the set $M^k = A^{(1)} \sim D_k$ and the function

$$g_k(y) = \int \chi_{\{f > y\}} \chi_{M^k} d\mathcal{L}^n.$$

As we showed in Step 1, for all sufficiently large k , the derivative of g_k will have an absolutely continuous piece in the interval (Y, Z) which is bounded below by some fixed nonzero function $\tilde{h}$. Therefore, for all sufficiently large k and $B \subset (Y^*, Z^*)$, we have

$$\int \chi_{\{f \in B\}} \chi_{\{W_j > 1/2\}} d\mathcal{L}^n \geq \int_B \tilde{h} d\mathcal{L}^1$$

with

$$\int_{Y^*}^{Z^*} \tilde{h} d\mathcal{L}^1 > 0.$$

Applying this inequality to $\nu_k(y)$, we have for $y \in I^k$ and $j = 2^k J$,

$$\nu_k(y) \geq \int_{y-C_5/j}^y \tilde{h}(y) d\mathcal{L}^1 y.$$

By the Schwarz inequality,

$$\nu_k(y) \geq (j^{-1}C_5)^{-1} \left[\int_{y-C_5/j}^y \tilde{h}(t)^{1/2} d\mathcal{L}^1 t \right]^2.$$

Therefore,

$$\Delta_k \geq (JC_5)^{-1/2} j \int_{Y^*}^{Z^*} \chi_k(y) \left(\int_{y-C_5/j}^y \tilde{h}(t)^{1/2} d\mathcal{L}^1 t \right) d\mathcal{L}^1 y.$$

We can change the order of integration so that

$$\Delta_k \geq (JC_5)^{-1/2} \int_{Y^*}^{Z^*} \tilde{h}(t)^{1/2} \sigma_k(t) d\mathcal{L}^1 t$$

with

$$\sigma_k(t) = j \int_t^{t+C_5/j} \chi_k(y) d\mathcal{L}^1 y.$$

Recalling the definition of χ_k as the characteristic function of I^k , it is easy to see that $\sigma_k(t)$ is bounded above by C_5 and that the measures $\mathcal{L}^1 \wedge \sigma_k$ converge weakly to the constant measure $\mathcal{L}^1 \wedge C_5/4$ on (Y^*, Z^*) as $k \rightarrow \infty$. This implies (since $\tilde{h}$ is a fixed function and the continuous functions are dense in $L^1(\mathbf{R})$) that

$$\limsup_{k \rightarrow \infty} \Delta_k \geq (JC_5)^{-1/2} (C_5/4) \int_{Y^*}^{Z^*} \tilde{h}(t)^{1/2} d\mathcal{L}^1 t > 0,$$

which was our goal. ■

Remark. In our proof above, the failure of convergence of ∇f_j^* to ∇f^* was ultimately related to an integral of $h^{1/2}$, where the function h is the absolutely continuous part of the derivative of $\mathcal{G}_f$. This is satisfying from the following point of view. We proved elsewhere that the singular part of the derivative of $\mathcal{G}_f$ cannot prevent the convergence of ∇f_j^* to ∇f^* , and this fact shows up clearly in our proof because the “square root of a singular measure” is zero (by this we mean that if one approximates a singular measure by an absolutely continuous one, then the square root converges to zero).

Theorem 4.3 states that the co-area irregularity of f always implies the existence of a sequence f_j which converges to f in $\mathbf{W}^{1,1}(\mathbf{R}^n)$ but for which f_j^* does not converge for f^* in $\mathbf{W}^{1,1}(\mathbf{R}^n)$. This sequence f_j is complicated, and it might be supposed that imposing some restriction on the f_j ’s would restore the convergence of f_j^* to f^* . The strongest condition one might think of imposing is that $f_j = f + t_j g$, with $t_j \in \mathbf{R}^+$ tending to zero and with $g \in C_c^\infty(\mathbf{R}^n)$. The following theorem shows that even such a strong supplementary condition does not, in general, imply that f_j^* converges to f^* .

6.2. Theorem (Rearrangement need not be continuous even for one-dimensional affine perturbations). *Let $n \geq 2$ and let f be one of the co-area irregular $C^{n-1,\lambda}(\mathbf{R}^n)$ functions constructed in 5.1. No condition is imposed on $\lambda \in (0, 1)$ but we require that $0 < \alpha < \frac{1}{4}$. Let $g : \mathbf{R}^n \rightarrow \mathbf{R}^+$ be any $C_c^\infty(\mathbf{R}^n)$ function with the property that*

$$g(x) = g(x_1, x_2, \dots, x_n) = a - x_1 \quad \text{for some } a > 0$$

whenever $x \in Q_0 = \{z : |z_i| \leq 1 \text{ for each } i = 1, \dots, n\}$. Then there is a sequence of positive numbers $t_1, t_2, t_3, \dots$ tending to zero (and which depends only on f and not on g) such that the sequence of functions

$$f_j \equiv f + t_j g$$

(which converges to f in $\mathbf{W}^{1,p}(\mathbf{R}^n)$ for all p) has the property that the f_j^ ’s do not converge to f^* in $\mathbf{W}^{1,1}(\mathbf{R}^n)$ (and hence also do not converge in $\mathbf{W}^{1,p}(\mathbf{R}^n)$).*

Proof. The notation and details of the construction in 5.1 will be freely used here. We define the t_J 's for $J \geq 1$ by setting

$$t_J = \frac{3^{-nJ}}{2L_J} \leq \frac{B}{2} 3^{-(n-1)J}.$$

Let us study $f_J(x) = f(x) + t_J g$ for $x \in S_\infty$ which, apart from a set of hyperplanes of zero $\mathcal{L}^n$ measure, is the set on which $\nabla f = 0$. Recall there is a unique infinite sequence $\sigma = \{\sigma(1), \sigma(2), \dots\}$ for each such x such that

$$(1) \quad \{x\} = \bigcap_{j=1}^{\infty} Q_j(\sigma) = \bigcap_{j=1}^{\infty} P_j(\sigma) \quad \text{and} \quad f(x) = \sum_{j=1}^{\infty} N_j(\sigma) 3^{-nj}.$$

Thus, there is a one-to-one correspondence between S_∞ and Σ , the set of σ 's, and formula (1) gives all possible values of $f(x)$ as x ranges over S_∞ . Unfortunately, there is no very simple formula for x in terms of σ .

Note that for each J and σ , the point $2 \sum_{k=1}^J L_k \sigma(k) \equiv C_J(\sigma) \in Q_0$, which is the center of the cube $Q_J(\sigma)$, is in S_∞ . The reason is that if we define

$$\tilde{\sigma} = \{\sigma(1), \sigma(2), \dots, \sigma(J), 0, 0, \dots\},$$

then from the definition of $C_j(\sigma)$, we find that $C_j(\tilde{\sigma}) = C_j(\sigma)$ for all $j \geq J$ and thus $C_J(\sigma) \in \bigcap_{j=1}^{\infty} Q_j(\tilde{\sigma}) \subset S_\infty$. Now fix σ and J , and consider the special point $C_J(\sigma) \in S_\infty$. (*Caution.* The sequence corresponding to $C_j(\sigma)$ is not necessarily σ itself; it is the sequence $\tilde{\sigma}$.) Let us compute $f_J(C_J(\sigma)) = f(C_J(\sigma)) + t_J(C_J(\sigma))_1$. Recalling that $N_j(\tilde{\sigma}) = \sum_{s=1}^n [\tilde{\sigma}(j)_s + 1] 3^{s-1}$, we have

$$\begin{aligned} f_J(C_J(\sigma)) &= at_J + \sum_{\substack{j=1 \\ j \neq J}}^{\infty} 3^{-nJ} N_j(\tilde{\sigma}) \\ &\quad + 3^{-nJ} \left[\sum_{s=1}^n [\sigma(J)_s + 1] 3^{s-1} - \frac{2}{2L_J} \sum_{k=1}^J L_k \sigma(k)_1 \right] \end{aligned}$$

from which we see that $f_J(C_J(\sigma))$ does not depend on $\sigma(J)_1$ because of cancellation in the last summand on the right. Furthermore, since $\tilde{\sigma}(j) = 0$ for $j > J$, it is also true that $f_j(C_J(\sigma))$ is independent of $\sigma(J+1)$, $\sigma(J+2)$, $\dots$.

Next, with σ and J still fixed, let x range over the cube $Q_J(\sigma)$, but with $x \in S_\infty$. This means that we consider all possible sequences σ^* with the property that $\sigma^*(j) = \sigma(j)$ for $j = 1, \dots, J$. Since $N_j(\sigma^*) \in \{0, 1, \dots, 3^n - 1\}$ for each j and $N_j(\tilde{\sigma}) = \frac{1}{2}(3^n - 1)$ for $j > J$, we have that $|N_j(\sigma^*) - N_j(\tilde{\sigma})| \leq \frac{1}{2}(3^n - 1)$ when $j > J$, and hence from (1) that

$$(2) \quad |f(x) - f_J(C_J(\sigma))| \leq \sum_{j=J+1}^{\infty} \frac{1}{2} (3^n - 1) 3^{-nj} = \frac{1}{2} 3^{-nJ}.$$

Likewise, the perturbation $t_J g$ satisfies

$$(3) \quad |t_J g(x) - t_J g(C_J(\sigma))| = t_J |(x - C_J(\sigma))_1| \leq t_J L_J = \frac{1}{2} 3^{-nJ},$$

and thus for $x \in Q_J(\sigma) \cap S_\infty$,

$$(4) \quad |f_J(x) - f_J(C_J(\sigma))| \leq 3^{-n_J}.$$

With J fixed, there are 3^{n_J} distinct (and disjoint) cubes $Q_J(\sigma)$ with $\sigma \in \Sigma$. Let W_J denote the union of these 3^{n_J} cubes, so that $S_\infty = W_J \cap S_\infty$ and

$$(5) \quad f(S_\infty) = f(W_J \cap S_\infty) = [0, 1].$$

Equation (5) can be interpreted as follows (the remarks in this paragraph are for clarification only, and are not needed in our proof). It is easy to check that as $C_J(\sigma)$ runs over the 3^{n_J} centers of the cubes, $f(C_J(\sigma))$ takes on the values $A, 3A, 5A, \dots, (2 \cdot 3^{n_J} - 1)A$, where $A \equiv \frac{1}{2}3^{-n_J}$. Furthermore, inequality (2) is, in fact, a correct assessment of the range of $f(x)$ in each cube for $x \in S_\infty$. In other words, the range of $f(x)$ with $x \in S_\infty$ in the 3^{n_J} cubes is $[0, 2A], [2A, 4A], \dots, [2(3^{n_J} - 1)A, 2 \cdot 3^{n_J}A]$, and this is how the range $[0, 1]$ is achieved.

Now consider the function f_J restricted to S_∞ . Inequality (4) states that the range of f_J in each cube can be larger than the previous value $2A$, but it can be at most $4A$. The crucial point, however, and the reason we used a triadic decomposition in 5.1, is that $f_J(C_J(\sigma))$ does not depend on $\sigma(J)_1$ and does not depend on $\sigma(j)$ for $j > J$. This means that each of the numbers $f_J(C_J(\sigma))$ is taken with multiplicity at least three so that only at most $\frac{1}{3}3^{n_J}$ distinct values are taken instead of the 3^{n_J} distinct values taken on by $f(C_J(\sigma))$. This implies for every J that

$$(6) \quad \mathcal{L}^1(f_J(S_\infty)) \leq \frac{2}{3},$$

which is to be compared with $\mathcal{L}^1(f(S_\infty)) = 1$. We do not assert that $f_J(S_\infty) \subset [0, 1]$.

Our strategy for proving the failure of $\mathbf{W}^{1,1}(\mathbf{R}^n)$ convergence of f_J^* to f^* will be to show that the necessary criterion (3) of Theorem 4.1 (which is proved in 8.8) is not satisfied. Recalling the definitions of the distribution functions $\mathcal{V}_f$, $\mathcal{F}_f$, and $\mathcal{G}_f$ in Definition 1.2.5, we shall use a subscript J to denote these distribution functions for f_J and use no subscript to denote these distribution functions for f . A superscript 1 or ∞ will denote the contributions to these distribution functions from S_1 or S_∞ . For example,

$$\mathcal{V}_J^\infty = \int_{S_\infty} \chi_{\{f_J > y\}} d\mathcal{L}^n.$$

From 5.1, we know that $\mathcal{G}(y) = \mathcal{G}^\infty(y) = C(1 - y)$ for $0 \leq y \leq 1$ with $C = 2^n(1 - \alpha)$. This $\mathcal{G}$ is absolutely continuous.

On the other hand, $\mathcal{F} = \mathcal{F}^1$ is also absolutely continuous (Theorem 3.2). Thus, in the notation of Theorem 4.1, the classical (or absolutely continuous) derivative $\mathcal{W}$ satisfies the condition

$$|\mathcal{W}(y)| \geq C = 2^n(1 - \alpha)$$

for almost every $0 \leq y \leq 1$.

Now consider the functions f_J and the classical derivatives $\mathscr{W}_J$ of $\mathscr{V}_J$. Should condition (3) of Theorem 4.1 be satisfied, there would be a subsequence $J(1), J(2), \dots$ such that

$$(7) \quad \liminf_{k \rightarrow \infty} |\mathscr{W}_{J(k)}(y)| \geq C$$

for $\mathscr{L}^1$ almost every y in $[0, 1]$. We shall show that (7) fails in $[\varepsilon, 1]$ for every small $\varepsilon > 0$, so that *a fortiori* it fails in $[0, 1]$. Choose $\varepsilon > 0$.

The values of $f_J(x)$ for $x \in S_0$ are $t_J g(x)$, and since g is bounded, these values do not exceed ε when J is large enough. Hence, the region S_0 can be ignored by ignoring small J 's henceforth.

The closure of the range of $f(x)$ for $x \in S_1$ is $[0, 1]$, and hence S_1 cannot be ignored. Recall, however, that in terms of the α in 5.1, we have $\mathscr{L}^n(S_1) = 2^n \alpha$ and $\mathscr{L}^n(S_\infty) = 2^n(1 - \alpha)$. When $\alpha < \frac{1}{4}$ and ε is small enough, S_1 can be made harmless, as will be shown.

For any Borel subset A of $[0, \infty)$, the derivative measures satisfy the condition

$$|\mathscr{V}_J^{\infty'}(A)| = \int_{S_\infty} \chi_{\{f_J \in A\}} d\mathscr{L}^n$$

and similarly for $\mathscr{V}_J^{1'}$. To prove failure of (7), an upper bound for $|\mathscr{W}_J|$ is needed. This can be done by supposing that the measures $\mathscr{V}_J^{\infty'}$ and $\mathscr{V}_J^{1'}$ have no singular parts. Thus,

$$\begin{aligned} \int_{[0, 1]} |\mathscr{W}_J^\infty| d\mathscr{L}^1 &\leq \int_{S_\infty} d\mathscr{L}^n = 2^n(1 - \alpha), \\ \int_{[0, 1]} |\mathscr{W}_J^1| d\mathscr{L}^1 &\leq \int_{S_1} d\mathscr{L}^n = 2^n \alpha. \end{aligned}$$

Clearly, $|\mathscr{W}_J(y)| = |\mathscr{W}_J^1(y)| + |\mathscr{W}_J^\infty(y)|$ when $y > \varepsilon$. By (6), there is a Borel subset B_J of $[0, 1]$ such that $\mathscr{L}^1(B_J) \geq \frac{1}{3}$ and $B_J \cap f(S_\infty)$ is empty. Therefore,

$$\int_{B_J} |\mathscr{W}_J^\infty| d\mathscr{L}^1 \leq |\mathscr{V}_J^{\infty'}(B_J)| = 0,$$

and hence $\mathscr{W}_J^\infty(y) = 0$ for almost every $y \in B_J$.

For $y \in [\varepsilon, 1]$, define

$$\varphi(y) \equiv \liminf_{k \rightarrow \infty} |\mathscr{W}_{J(k)}(y)|$$

and $R_k(y) \equiv \min\{\varphi(y), |\mathscr{W}_{J(k)}(y)|\}$. Clearly, $R_k(y) \rightarrow \varphi(y)$ as $k \rightarrow \infty$ for every y in $[\varepsilon, 1]$. By Egoroff's theorem, there is a measurable set $D \subset [\varepsilon, 1]$ with $\mathscr{L}^1(D) > 1 - 2\varepsilon$ such that R_k converges uniformly on D . This means that when k is large enough (which is henceforth assumed) $|\mathscr{W}_{J(k)}(y)| \geq \varphi(y) - \varepsilon$ for every y in D .

Suppose now that (7) is true for $\mathcal{L}^1$ almost every y in $[\varepsilon, 1]$ so that $\varphi(y) \geq C$. Then since $\mathcal{W}_J^\infty(y) = 0$ for every $y \in B_J$, we must have that $|\mathcal{W}_{J(k)}^1(y)| \geq C - \varepsilon$ for almost every y in $D \cap B_{J(k)}$. However,

$$\begin{aligned} 2^n \alpha &\geq \int_{[\varepsilon, 1]} |\mathcal{W}_{J(k)}^1| d\mathcal{L}^1 \geq \int_{D \cap B_{J(k)}} |\mathcal{W}_{J(k)}^1| d\mathcal{L}^1 \\ &\geq (C - \varepsilon) \mathcal{L}^1(D \cap B_{J(k)}) \\ &\geq (C - \varepsilon) \left(\frac{1}{3} - 2\varepsilon\right) \geq 2^n (1 - \alpha - \varepsilon) \left(\frac{1}{3} - 2\varepsilon\right). \end{aligned}$$

This implies that

$$\alpha \left(\frac{4}{3} - 2\varepsilon\right) \geq (1 - \varepsilon) \left(\frac{1}{3} - 2\varepsilon\right).$$

Since we assumed that $\alpha < \frac{1}{4}$, this last inequality can be violated by choosing ε small enough (depending only on α), and a contradiction with (7) has been established. ■

Remark. Theorem 6.2 has a geometric interpretation in $\mathbf{R}^{n+1}$ which is worth noting. Since g is affine on Q_0 , which is the support of f , the subgraph of the function $f_j = f + t_j g$ (which is $\{(x, y) : 0 \leq y \leq f(x)\}$) in the region $Q_0 \times [\varepsilon, \infty)$ is obtained from the subgraph A of f for large j 's by vertically translating A a very small amount and then very slightly shearing the translated set. It is indeed surprising that while the translated sheared subgraphs converge quite rapidly and strongly to the unperturbed subgraph, the boundary area of the symmetrization of this translated sheared subgraph does not converge to the boundary area of the symmetrization of the unperturbed subgraph. The justification for this latter assertion is Theorem 8.6 applied to the arc length function.

§7. CONVERGENCE WITH GENERAL INTEGRANDS IS EFFECTIVELY EQUIVALENT TO $\mathbf{W}^{1,1}$ CONVERGENCE

In this section, we want to relate the question of the convergence of integrals $\int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n$ to zero to the question of the convergence of $\|\nabla f_j^* - \nabla f^*\|_{L^1}$ to zero, and to show that these are effectively the same problem. In the next section, $\mathbf{W}^{1,1}$ convergence will be shown to be equivalent to arc length convergence, and this fact will lead immediately to the proof of Theorem 4.1. We begin here by showing the relationship in the easy direction, i.e., if $\int \Psi(|\nabla f_j^{*I} - \nabla f^{*I}|) d\mathcal{L}^n \rightarrow 0$, then for every strictly positive interval I , $\|\nabla f_j^{*I} - \nabla f^{*I}\|_{L^1} \rightarrow 0$; here the superscript I means that we consider only functional values in the interval I .

According to Theorem 2.7, the assumption that f belongs to $\mathcal{E}_1$ implies that f^* belongs to $\mathcal{E}_1$. Similarly, according to Theorem 2.6, the assumption that f belongs to $\mathcal{E}_2$ implies that f^* belongs to $\mathcal{E}_2$. For the purposes of Theorem 7.2 only, we assume that our integrand Ψ satisfies only the following minimal conditions:

- (1) $\Psi: \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is a Borel function with $\Psi(0) = 0$. We do not assume that Ψ is convex or continuous.

- (2) For each $\varepsilon > 0$, there is a positive constant C_ε such that $\Psi(t) \geq C_\varepsilon$ whenever $t \geq \varepsilon$.

Suppose now that $f^*, f_1^*, f_2^*, f_3^*, \dots : \mathbf{R}^n \rightarrow \mathbf{R}^+$ are nonnegative, symmetric decreasing functions which belong to $\mathcal{E}_1$ and that $\int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \rightarrow 0$ as $j \rightarrow \infty$. One is tempted to conclude from this that $|\nabla f_j^* - \nabla f^*| \rightarrow 0$ in $\mathbf{L}^1(\mathbf{R}^n)$, but it is easy to see that this conclusion cannot be valid generally (assuming only conditions (1) and (2)) because the f_j^* 's might have a decay at infinity that is too slow. What is true, provided an *extra assumption* is made, is that the gradients of the *slab functions* (defined below) converge in $\mathbf{L}^1(\mathbf{R}^n)$ when $Y > 0$.

The extra assumption will be that there is some fixed positive radius R such that $\lim_{j \rightarrow \infty} f_j^*(R) = f^*(R)$ (recall our conventions in Definition 1.3). This assumption will not rule out any interesting cases because it will automatically be satisfied in such cases. For example, if $f_j \rightarrow f$ in $\mathbf{W}^{1,p}(\mathbf{R}^n)$, then Remark (2) following Corollary 2.3 implies that $f_j^* \rightarrow f^*$ in $\mathbf{L}^p(\mathbf{R}^n)$. If the ∇f_j^{*I} 's did not converge to ∇f^{*I} in $\mathbf{L}^1(\mathbf{R}^n)$, there would exist a positive number γ together with a subsequence $j(k)$ of the j 's for which $\|\nabla f_{j(k)}^{*I} - \nabla f^{*I}\|_{\mathbf{L}^1} > \gamma$ for each k . Passing to a further subsequence if necessary, we can assume that the $f_{j(k)}^*$'s converge to f^* pointwise $\mathcal{L}^n$ almost everywhere. For this subsequence, our extra assumption above is certainly satisfied. The following Theorem 7.2 would then contradict the assumed failure of convergence. An alternative hypothesis that would allow the same conclusion (by virtually the same reasoning) is that the f_j 's converge to f in measure.

7.1. Definition. Let $0 \leq Y < Z \leq \infty$, and set $I = (Y, Z) \subset \mathbf{R}^+$. Associated with any function $f : \mathbf{R}^n \rightarrow \mathbf{R}^+$ is its I -slab function f^I defined by setting

$$f^I(x) = \begin{cases} 0 & \text{in case } f(x) \leq Y, \\ f(x) - Y & \text{in case } Y \leq f(x) \leq Z, \\ Z - Y & \text{in case } Z \leq f(x). \end{cases}$$

Note that $(f^*)^I$ equals $(f^I)^*$ whenever $f : \mathbf{R}^n \rightarrow \mathbf{R}^+$. Another fact that is obvious from the definition is that f^I is co-area regular if f is co-area regular.

7.2. Theorem (Ψ convergence effectively implies $\mathbf{W}^{1,1}$ convergence). Suppose $\Psi : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is an integrand that satisfies conditions (1) and (2) above. Let $f^*, f_1^*, f_2^*, \dots : \mathbf{R}^n \rightarrow \mathbf{R}^+$ be nonnegative, symmetric decreasing functions which belong to our class $\mathcal{E}_1$. Suppose also

- (a) $\int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \rightarrow 0$ as $j \rightarrow \infty$;
and

- (b) $\lim_{j \rightarrow \infty} f_j^*(R) = f^*(R)$ for some fixed positive radius R .
Then

- (1) for each $r > 0$, $f_j^*(r) \rightarrow f^*(r)$ as $j \rightarrow \infty$.
(2) Whenever $0 < Y < Z \leq \infty$ and $I = (Y, Z)$, the I -slab functions f_j^{*I} converge to the I -slab function f^{*I} strongly in $\mathbf{W}^{1,1}(\mathbf{R}^n)$.

Proof. First part. Since f^* is radial, we can follow the notational conventions given in Definition 1.3 and further abuse notation to abbreviate $r = |x|$ so that $\nabla f^*(r) = (df^*/dr)(r) \leq 0$ has the obvious meaning. For each r , we can use the fundamental theorem of calculus for distributional derivatives (which in the present case are $\mathbf{L}_{\text{loc}}^1(\mathbf{R}^n)$ functions) to write

$$f_j^*(r) = f_j^*(R) - \int_r^R \nabla f_j^* d\mathcal{L}^1.$$

Setting $A_j(\varepsilon) = \{x : |\nabla f_j^*(x) - \nabla f^*(x)| > \varepsilon\}$ for small positive numbers ε , we can rewrite the integral on the right side above to become

$$\int_r^R \chi_{A_j(\varepsilon)} \nabla f_j^* d\mathcal{L}^1 + \int_r^R (1 - \chi_{A_j(\varepsilon)}) \nabla f_j^* d\mathcal{L}^1.$$

For $x \in A_j(\varepsilon)$, condition (2) above on Ψ implies that

$$|\nabla f_j^* - \nabla f^*| \leq \Psi(|\nabla f_j^* - \nabla f^*|)/C_\varepsilon,$$

while, of course, for $x \notin A_j(\varepsilon)$ we have $|\nabla f_j^* - \nabla f^*| \leq \varepsilon$. Since $f^*(r)$ satisfies the corresponding equations, we have

$$\begin{aligned} |f_j^*(r) - f^*(r)| &\leq |f_j^*(R) - f^*(R)| \\ &\quad + \frac{1}{C_\varepsilon} \int \chi_{\{s \geq \min\{r, R\}\}} \Psi(|\nabla f_j^*(s) - \nabla f^*(s)|) d\mathcal{L}^1 s + \varepsilon |R - r|. \end{aligned}$$

We then use polar coordinates to write

$$\begin{aligned} &\int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \\ &\geq \int \chi_{\{|x| \geq \min\{r, R\}\}} \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \\ &= n\alpha(n) \int \chi_{\{s \geq \min\{r, R\}\}} \Psi(|\nabla f_j^*(s) - \nabla f^*(s)|) s^{n-1} d\mathcal{L}^1 s \\ &\geq n\alpha(n) \min\{r, R\}^{n-1} \int \chi_{\{s \geq \min\{r, R\}\}} \Psi(|\nabla f_j^*(s) - \nabla f^*(s)|) d\mathcal{L}^1 s; \end{aligned}$$

here $\alpha(n)$ is the volume of the unit ball in $\mathbf{R}^n$. Letting $j \rightarrow \infty$, we conclude $\lim_{j \rightarrow \infty} |f_j^*(r) - f^*(r)| \leq \varepsilon |R - r|$. We finish the proof of conclusion (1) by letting $\varepsilon \rightarrow 0$.

Second part. For the following reason, it suffices to prove our second conclusion in the case that $Z = \infty$. Assuming the result in this case, we could in the general case set $J = (Z, \infty)$ and $K = (Y, \infty)$ so that (up to sets of measure zero) $\nabla f^{*I} = \nabla f^{*K} - \nabla f^{*J}$. Then

$$|\nabla f_j^{*I} - \nabla f^{*I}| \leq |\nabla f_j^{*K} - \nabla f^{*K}| + |\nabla f_j^{*J} - \nabla f^{*J}|,$$

and both terms on the right side converge to zero in $\mathbf{L}^1(\mathbf{R}^n)$.

Henceforth, we set $Z = \infty$.

Since f^* is monotone nonincreasing and belongs to $\mathcal{E}_1$, there is a positive finite radius R such that $f^*(x) < Y$ whenever $|x| \geq R$; we set $D = \{x : |x| < R\}$, noting, in particular, that $\mathcal{L}^n(D) < \infty$. Since the f_j^* 's converge to f^*

pointwise (by the first part), the supports of f^{*I} and all the f_j^{*I} 's will lie in D for all large j 's. We first assert that

$$(1) \quad \int \chi_D |\nabla f_j^* - \nabla f^*| d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

We set $\varepsilon = 1$ in the definition of our sets $A_j(\varepsilon)$ given above and abbreviate $A_j = A_j(1)$. On A_j , $|\nabla f_j^* - \nabla f^*| \leq \Psi(|\nabla f_j^* - \nabla f^*|)/C_1$. We also note that since $\Psi(|\nabla f_j^* - \nabla f^*|) \rightarrow 0$ in $L^1(\mathbf{R}^n)$, we can pass to a subsequence such that, $\mathcal{L}^n$ almost everywhere, $\Psi(|\nabla f_j^* - \nabla f^*|) \rightarrow 0$. Condition (2) on Ψ implies that, $\mathcal{L}^n$ almost everywhere, $|\nabla f_j^* - \nabla f^*| \rightarrow 0$. Then

$$\int \chi_D \chi_{A_j} |\nabla f_j^* - \nabla f^*| d\mathcal{L}^n \leq \frac{1}{C_1} \int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \rightarrow 0$$

and, on the other hand,

$$\int \chi_D (1 - \chi_{A_j}) |\nabla f_j^* - \nabla f^*| d\mathcal{L}^n \rightarrow 0$$

by Lebesgue's dominated convergence theorem (since the integrand is bounded by $\chi_D \in L^1(\mathbf{R}^n)$). Thus, (1) is proved.

Now we write

$$\|\nabla f_j^{*I} - \nabla f^{*I}\|_{L^1} = L_j + M_j + N_j$$

with

$$\begin{aligned} L_j &= \int \chi_D \chi_{\{f^* > Y\}} \chi_{\{f_j^* > Y\}} |\nabla f_j^* - \nabla f^*| d\mathcal{L}^n, \\ M_j &= \int \chi_D \chi_{\{f^* \leq Y\}} \chi_{\{f_j^* > Y\}} |\nabla f_j^*| d\mathcal{L}^n, \\ N_j &= \int \chi_D \chi_{\{f^* > Y\}} \chi_{\{f_j^* \leq Y\}} |\nabla f^*| d\mathcal{L}^n. \end{aligned}$$

Conclusion (2) of the present theorem is the assertion that $L_j + M_j + N_j \rightarrow 0$ as $j \rightarrow \infty$. The convergence in (1) above implies that $L_j \rightarrow 0$ as $j \rightarrow \infty$. Suppose then that $M_j + N_j \not\rightarrow 0$. The convergence in (1) implies that we can fix a function G in $L^1(\mathbf{R}^n)$ and a subsequence of the j 's (without changing notation) such that $|\nabla f_j| \leq G$ for each j and $|\nabla f_j^* - \nabla f^*| \rightarrow 0$ almost everywhere on D . Since $\nabla f^*(x) = 0$ at almost every x for which $f^*(x) = Y$, we use Lebesgue's dominated convergence theorem to infer

$$\int \chi_D \chi_{\{f^* = Y\}} \chi_{\{f_j^* > Y\}} |\nabla f_j| d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty.$$

We note also that

$$\int \chi_D \chi_{\{f^* < Y\}} \chi_{\{f_j^* > Y\}} |\nabla f_j| d\mathcal{L}^n \leq \int \chi_D \chi_{\{f^* < Y\}} \chi_{\{f_j^* > Y\}} G d\mathcal{L}^n \rightarrow 0 \quad \text{as } j \rightarrow \infty$$

by virtue of Lebesgue's dominated convergence theorem again, since the fact that $f_j^* \rightarrow f^*$ almost everywhere implies $\chi_{\{f^* < Y\}} \chi_{\{f_j^* > Y\}} \rightarrow 0$ almost everywhere. Hence, $M_j \rightarrow 0$ as $j \rightarrow \infty$. An argument virtually identical to (2) shows $N_j \rightarrow 0$ also. This contradiction finishes the proof of (2). ■

Our next goal is to prove that $\mathbf{W}^{1,1}$ convergence effectively implies convergence of $\int \Psi(|\nabla f_j - \nabla f|) d\mathcal{L}^n$ to zero for suitable convex Ψ . The main theorem here is Theorem 7.5; it will be used in the next part. We begin by proving convergence of slab functions f^I . Note that the convex integrand Ψ in Theorem 7.3 is more general than the spherically symmetric Ψ of Theorems 4.1 and 4.2. The hypothesis that each subsequence of the f_j 's have a further subsequence that converges to f almost everywhere is rather mild; it is guaranteed, for instance, by the hypothesis that $f_j \rightarrow f$ in $L^p(\mathbf{R}^n)$ or that $f_j \rightarrow f$ in measure as $j \rightarrow \infty$.

7.3. Theorem (General Ψ convergence of slab functions). *Suppose that $\Psi: \mathbf{R}^n \rightarrow \mathbf{R}^+$ is convex and satisfies the conditions that $\Psi(0) = 0$, $\Psi(-\xi) = \Psi(\xi)$, and $\Psi(\xi) \leq C\Psi(\eta)$ for some fixed $C \geq 1$ whenever $\xi, \eta \in \mathbf{R}^n$ with $|\xi| \leq 2|\eta|$. Suppose also that the functions $f, f_1, f_2, f_3, \dots: \mathbf{R}^n \rightarrow \mathbf{R}^+$ belong to our class $\mathcal{E}_1$ and satisfy the additional condition that each subsequence of the f_j 's has a further subsequence whose members converge to f pointwise almost everywhere. Finally, assume that $\Psi(\nabla f), \Psi(\nabla f_1), \Psi(\nabla f_2), \Psi(\nabla f_3), \dots$ are all summable and that $\Psi(\nabla f_j - \nabla f) \rightarrow 0$ in $L^1(\mathbf{R}^n)$ as $j \rightarrow \infty$. Then whenever $0 \leq Y < Z \leq \infty$ and $I = (Y, Z)$, we have the following convergence in $L^1(\mathbf{R}^n)$ as $j \rightarrow \infty$ (recall the definition of slab functions in Definition 7.1).*

- (1) $\Psi(\nabla f_j) \rightarrow \Psi(\nabla f)$.
- (2) $\Psi(\nabla f_j^I - \nabla f^I) \rightarrow 0$.
- (3) $\Psi(\nabla f_j^I) \rightarrow \Psi(\nabla f^I)$.

Proof. We shall use the following inequality for convex functions, which is Lemma 3 in [BL]. Suppose $\xi, \eta \in \mathbf{R}^n$, $0 < \varepsilon < \frac{1}{2}$, and $\Phi: \mathbf{R}^n \rightarrow \mathbf{R}$ is convex (but not necessarily nonnegative). Then

$$|\Phi(\xi + \eta) - \Phi(\xi)| \leq \varepsilon[\Phi(2\xi) - 2\Phi(\xi)] + |\Phi(\frac{\eta}{\varepsilon})| + |\Phi(-\frac{\eta}{\varepsilon})|.$$

In our case (with Φ replaced by Ψ), the last two terms are identical since $\Psi(-\xi) = \Psi(\xi)$.

Step 1. We will show that (1) is a consequence of our assumption that $\Psi(\nabla f_j^I - \nabla f^I) \rightarrow 0$ in $L^1(\mathbf{R}^n)$. This then implies that (3) is a consequence of (2). We set $\xi = \nabla f(x)$ and $\eta = \nabla f_j(x) - \nabla f(x)$ and use our growth condition on Ψ to estimate

$$\begin{aligned} |\Psi(\nabla f_j) - \Psi(\nabla f)| &\leq \varepsilon[\Psi(2\nabla f) - 2\Psi(\nabla f)] + 2\Psi(\frac{1}{\varepsilon}(\nabla f_j - \nabla f)) \\ &\leq \varepsilon C\Psi(\nabla f) + 2C^{2-\ln \varepsilon / \ln 2}\Psi(\nabla f_j - \nabla f). \end{aligned}$$

To obtain our asserted convergence, we first fix $\varepsilon > 0$ and integrate our extreme inequality over $\mathbf{R}^n$ and then let $j \rightarrow \infty$. The second term on the right converges to zero. We then let $\varepsilon \rightarrow 0$ to establish (1).

Step 2. We will now prove (2). For simplicity, we consider only the case $Y = 0$ and $Z < \infty$. The proof for $0 < Y < Z = \infty$ is similar, and the proof for $0 < Y < Z < \infty$ is a consequence of these two cases. We note that

$\nabla f^j(x) = \nabla f(x)$ if $f(x) < Z$ while $\nabla f(x) = 0$ if $f(x) \geq Z$; recall that $\nabla f = 0$ almost everywhere on the set $\{x : f(x) = Z\}$. We set

$$A_j^+ = \{x : f(x) > Z \text{ and } f_j(x) < Z\}, \quad A_j^- = \{x : f(x) < Z \text{ and } f_j(x) > Z\}$$

and let $\chi_j^\pm$ be the respective characteristic functions. Then we check that

$$\begin{aligned} \int \Psi(\nabla f_j^l - \nabla f^l) d\mathcal{L}^n &\leq \int \Psi(\nabla f_j - \nabla f) d\mathcal{L}^n \\ &\quad + \int \chi_j^- \Psi(\nabla f) d\mathcal{L}^n + \int \chi_j^+ \Psi(\nabla f_j) d\mathcal{L}^n. \end{aligned}$$

If (1) should fail, there would be a subsequence of the j 's (still denoted j) such that $\int \Psi(\nabla f_j^l - \nabla f^l) d\mathcal{L}^n \geq \alpha > 0$ for each j while $f_j \rightarrow f$ almost everywhere as $j \rightarrow \infty$. The convergence of the f_j 's implies that $\chi_j^\pm \rightarrow 0$ almost everywhere as $j \rightarrow \infty$. The first and second integrals on the right side obviously converge to zero as $j \rightarrow \infty$. To estimate the third integral, we first use the triangle inequality and then use our convexity inequality above with $\xi = \nabla f$, $\eta = \nabla f_j - \nabla f$, and $\varepsilon = \frac{1}{4}$ to obtain

$$\begin{aligned} \Psi(\nabla f_j) &\leq \Psi(\nabla f) + |\Psi(\nabla f_j) - \Psi(\nabla f)| \\ &\leq \Psi(\nabla f) + \frac{1}{4}[\Psi(2\nabla f) - 2\Psi(\nabla f)] + 2\Psi(4(\nabla f_j - \nabla f)). \end{aligned}$$

For the integral we wish to estimate, we multiply the extreme inequality by χ_j^+ and integrate over $\mathbf{R}^n$. The rightmost term converges to zero as $j \rightarrow \infty$ because $0 \leq \chi_j^+ \leq 1$ and $\Psi(4\xi) \leq C^2\Psi(\xi)$. There remains χ_j^+ times a fixed $L^1(\mathbf{R}^n)$ function, namely, $g = (\frac{1}{4}C + \frac{1}{2})\Psi(\nabla f)$. However, $\chi_j^+ \rightarrow 0$ almost everywhere as noted above, so that (1) follows from Lebesgue's dominated convergence theorem. ■

7.4. Theorem ($W^{1,1}$ convergence implies convergence for convex integrands and $\mathcal{E}_2$ functions). Suppose that $\Psi : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is a suitable integrand (recall Definition 2.1) and that the functions $f, f_1, f_2, f_3, \dots : \mathbf{R}^n \rightarrow \mathbf{R}^+$ belong to our class $\mathcal{E}_2$. Assume also that these functions and their rearrangements have the following convergence properties in $L^1(\mathbf{R}^n)$ as $j \rightarrow \infty$:

- (a) $|\nabla f_j - \nabla f| \rightarrow 0$;
- (b) $|\nabla f_j^* - \nabla f^*| \rightarrow 0$;
- (c) $\Psi(|\nabla f_j - \nabla f|) \rightarrow 0$;

and that, also,

- (d) $\Psi(|\nabla f_j|)$ and $\Psi(|\nabla f|)$ are in $L^1(\mathbf{R}^n)$.

Then the following convergences also occur in $L^1(\mathbf{R}^n)$ as $j \rightarrow \infty$.

- (1) $\Psi(|\nabla f_j^* - \nabla f^*|) \rightarrow 0$.
- (2) $\Psi(|\nabla f_j^*|) \rightarrow \Psi(|\nabla f^*|)$.
- (3) $\Psi(|\nabla f_j|) \rightarrow \Psi(|\nabla f|)$.

Proof. According to Theorem 2.7, the functions ∇f^* and the ∇f_j^* 's belong to $L^1(\mathbf{R}^n, \mathbf{R}^n)$. According to Theorem 2.5, the functions f_j , f , f_j^* and f^* belong to $L^p(\mathbf{R}^n)$ with $p = n/(n-1)$ ($n > 1$) or $p = \infty$ ($n = 1$), and by hypotheses (a) and (b), $f_j \rightarrow f$ and $f_j^* \rightarrow f^*$ in $L^p(\mathbf{R}^n)$. According to Theorem 7.3, (3) follows from hypotheses (a), (c), and (d). Likewise (2) follows from hypothesis (b), and conclusion (1) and from Theorem 2.7, which states that hypothesis (d) implies that $\Psi(|\nabla f_j^*|)$ and $\Psi(|\nabla f^*|)$ are in $L^1(\mathbf{R}^n)$. Thus, we can concentrate on proving (1) under the assumption that hypotheses (b) and (d) hold for the rearranged functions f_j^* and f^* .

Assuming the failure of (1), we infer the existence of $\gamma > 0$ and a subsequence of the f_j 's (again denoted by f_j) such that

$$\|\Psi(|\nabla f_j^* - \nabla f^*|)\|_{L^1} \geq 8\gamma.$$

We shall derive a contradiction from this.

The first step is to truncate f and the f_j 's at a suitable height Y . The functions H and H_j of Theorem 3.3 (with Φ there replaced by Ψ) are all continuous and vanish at infinity so there is a $Y > 0$ and a $\beta > 0$ (assuming that f is not the zero function) such that

$$H(Y) \leq \gamma C^{-2} \quad \text{and} \quad \mathcal{L}^n\{x : f > Y\} = 2\beta > 0.$$

By (2) and (3) of Theorem 3.3, we can assume that for large enough j ,

$$H_j(Y) \leq 2\gamma C^{-2} \quad \text{and} \quad \mathcal{L}^n\{x : f_j > Y\} \geq \beta > 0.$$

(The right statement is a consequence of the $L^p(\mathbf{R}^n)$ convergence of f^j 's to f .) Now set $I = (0, Y)$, and let f^I and f_j^I denote the slab functions of Definition 7.1. Note that $f^{I*} = f^{*I}$ and that

$$\int \Psi(|\nabla f|) d\mathcal{L}^n = \int \Psi(|\nabla f^I|) d\mathcal{L}^n + \int \chi_{\{f>Y\}} \Psi(|\nabla f|) d\mathcal{L}^n;$$

similar equations hold for f^* , f_j , and f_j^* . According to Theorem 7.3, $|\nabla f_j^{I*} - \nabla f^{I*}| \rightarrow 0$ in $L^1(\mathbf{R}^n)$ as $j \rightarrow \infty$ since $|\nabla f_j^* - \nabla f^*| \rightarrow 0$ in $L^1(\mathbf{R}^n)$. Finally, by Theorem 7.3, $\|\Psi(|\nabla f_j^I - \nabla f^I|)\|_{L^1} \rightarrow 0$ as $j \rightarrow \infty$. We have thus shown that the functions f^I and f_j^I (and their rearrangements) satisfy the hypotheses of Theorem 7.4.

In the same manner, we can let $J = (Y, \infty)$ be the complementary interval and infer that the functions f^J and f_j^J also satisfy the hypotheses of Theorem 7.4. We note that $f = f^I + f^J$, $f_j = f_j^I + f_j^J$, $f^* = f^{I*} + f^{J*}$, etc. Since Ψ is convex, Ψ is monotone, whence,

$$\Psi(|v + w|) \leq \Psi(|v| + |w|) \leq \frac{1}{2}\Psi(2|v|) + \frac{1}{2}\Psi(2|w|).$$

Using this, we have

$$\begin{aligned}
 \Psi(|\nabla f_j^* - \nabla f^*|) &\leq \frac{1}{2}\Psi(2|\nabla f_j^{I*} - \nabla f^{I*}|) + \frac{1}{2}\Psi(2|\nabla f_j^{J*} - \nabla f^{J*}|) \\
 &\leq \frac{C}{2}\Psi(|\nabla f_j^{I*} - \nabla f^{I*}|) + \frac{C}{2}\Psi(|\nabla f_j^{J*} - \nabla f^{J*}|) \\
 (*) \qquad &\leq \frac{C}{2}\Psi(|\nabla f_j^{I*} - \nabla f^{I*}|) + \frac{C^2}{4}\Psi(|\nabla f_j^{J*}|) + \frac{C^2}{4}\Psi(|\nabla f^{J*}|).
 \end{aligned}$$

As a consequence of Theorem 2.7, we have

$$\begin{aligned}
 \int \Psi(|\nabla f^{J*}|) d\mathcal{L}^n &= \int \chi_{\{f^* > Y\}} \Psi(|\nabla f^*|) d\mathcal{L}^n \leq \int \Psi(|\nabla f^J|) d\mathcal{L}^n \\
 &= \int \chi_{\{f > Y\}} \Psi(|\nabla f|) d\mathcal{L}^n \\
 &= H(Y),
 \end{aligned}$$

Similarly, $\int \Psi(|\nabla f_j^{I*}|) d\mathcal{L}^n \leq H_j(Y)$.

Let us assume, provisionally, that (1) of the present theorem holds for f^{I*} and the f_j^{I*} 's. Then integrating the above inequality (*) over $\mathbf{R}^n$ and letting $j \rightarrow \infty$, we infer

$$\begin{aligned}
 8\gamma &\leq \liminf_{j \rightarrow \infty} \int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \\
 &\leq \frac{C^2}{4} \liminf_{j \rightarrow \infty} (H(Y) + H_j(Y)) \leq \frac{3}{4}\gamma
 \end{aligned}$$

in view of our assumptions about $H(Y)$ and the $H_j(Y)$'s. This is clearly not possible. Therefore, in order to finish the proof of our theorem, it suffices to show that f_j^{I*} and f^{I*} satisfy (1).

Concerning these functions, we note one *crucial point*. If we define $R > 0$ by setting $\alpha(n)R^n = \beta$, the conditions

$$\mathcal{L}^n\{x : f_j > Y\} \geq \beta > 0 \quad \text{and} \quad \mathcal{L}^n\{x : f > Y\} \geq \beta > 0$$

then imply that $f^{I*}(x) = Y$ and $f_j^{I*}(x) = Y$ whenever $0 \leq |x| < R$. Henceforth, we shall drop the superscript I and will prove (1) under the *additional assumption of the existence of these constants Y and R* .

Since $\nabla f_j^* \rightarrow \nabla f^*$ in $\mathbf{L}^1(\mathbf{R}^n, \mathbf{R}^n)$ and $f_j^* \rightarrow f^*$ in $\mathbf{L}^p(\mathbf{R}^n)$, there is a subsequence (which we continue to denote by f_j) and an $\mathbf{L}^1(\mathbf{R}^n)$ function $G : \mathbf{R}^n \rightarrow \mathbf{R}^+$ such that $|\nabla f_j^*| \leq G$, $|\nabla f^*| \leq G$, and, as $j \rightarrow \infty$, $\nabla f_j^* \rightarrow \nabla f^*$ almost everywhere, and $f_j^* \rightarrow f^*$ almost everywhere. Clearly, we can take G to be spherically symmetric with $G(x) = 0$ whenever $|x| < R$. For fixed $T > 0$, we set $A_T = \{x : G(x) > T\} \subset \mathbf{R}^n$.

The growth condition on Ψ in Definition 2.1 implies that for each $T > 0$, there is a constant C_T such that $\Psi(|\xi|) \leq C_T|\xi|$ when $|\xi| \leq 2T$. If $x \notin A_T$, then $|\nabla f_j^*(x) - \nabla f^*(x)| \leq 2T$ so that

$$\int_{\mathbf{R}_n \sim A_T} \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \leq C_T \int |\nabla f_j^* - \nabla f^*| d\mathcal{L}^n.$$

Since this converges to zero as $j \rightarrow \infty$, we can concentrate on bounding

$$\int_{A_T} \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n.$$

Since

$$2\Psi(|\xi - \eta|) \leq \Psi(|2\xi|) + \Psi(|2\eta|) \leq C[\Psi(|\xi|) + \Psi(|\eta|)],$$

it suffices to estimate

$$K(T) = \int_{A_T} \Psi(|\nabla f^*|) d\mathcal{L}^n \quad \text{and} \quad K_j(T) = \int_{A_T} \Psi(|\nabla f_j^*|) d\mathcal{L}^n.$$

We will conclude the proof of the present theorem by showing that for sufficiently large T , $CK(T) \leq \gamma$ and $CK_j(T) < \gamma$ for all j 's. Since $G \in L^1(\mathbf{R}^n)$, the Lebesgue measure of A_T converges to zero as $T \rightarrow \infty$. The regularity of Lebesgue measure guarantees the existence of a spherically symmetric open set $B_T \supset A_T$ such that $\mathcal{L}^n(B_T) \leq 2\mathcal{L}^n(A_T)$. This open set B_T is necessarily the union of a countable family of pairwise disjoint open annular regions and thus can be written

$$B_T = \bigcup_{i=1}^{\infty} \{x : \sigma_i(T) < |x| < \tau_i(T)\}$$

for suitable functions $\sigma_i(T) < \tau_i(T)$ ($i = 1, 2, 3, \dots$). Because of our additional assumption above, we can require that $\sigma_i(T) \geq \frac{R}{2}$ for each i . (This is where the crucial property enters that $\nabla f_j^*(x)$ and $\nabla f^*(x)$ all vanish when $|x| < R$.) We set $s_i(T) \equiv f^*(\sigma_i(T))$ and $t_i(T) \equiv f^*(\tau_i(T))$ for each i , and further define

$$E_T = \bigcup_{i=1}^{\infty} \{y : t_i(T) \leq y \leq s_i(T)\} \subset [0, Y].$$

Then $M(T) \equiv (f^*)^{-1}(A_T) \subset (f^*)^{-1}(B_T) \subset E_T$. According to the fundamental theorem of calculus for distributional derivatives,

$$\begin{aligned} \mathcal{L}^1(E_T) &= \sum_{i=1}^{\infty} [f^*(\sigma_i(T)) - f^*(\tau_i(T))] \\ &= - \sum_{i=1}^{\infty} \int_{\sigma_i(T)}^{\tau_i(T)} (\nabla f^*)(r) dr \leq \sum_{i=1}^{\infty} \int_{\sigma_i(T)}^{\tau_i(T)} G(r) dr. \end{aligned}$$

Since $\sigma_i(T) > \frac{R}{2}$ for each i , the right side of this inequality is bounded above by

$$Z(T) \equiv [n\alpha(n)]^{-1} \left(\frac{R}{2}\right)^{1-n} \int \chi_{B_T} G d\mathcal{L}^n.$$

Since $A_{T_1} \subset A_{T_2}$ if $T_1 > T_2$, we can assume that B_T has this same monotone property. Since $\mathcal{L}^n(B_T) \rightarrow 0$ as $T \rightarrow \infty$, we infer that $\chi_{B_T} \rightarrow 0$ almost everywhere as $T \rightarrow \infty$, and hence, by Lebesgue's dominated convergence theorem, that $\int \chi_{B_T} G d\mathcal{L}^n \rightarrow 0$ also. Therefore, given any $\delta > 0$, we can find a $T(\delta)$ such that

$$\mathcal{L}^1(M(T(\delta))) \leq Z(T(\delta)) \leq \delta.$$

We note further that this estimate in terms of $Z(T)$ did not depend on f except insofar as $|\nabla f^*| \leq G$. Exactly the same estimate holds for f_j because $|\nabla f_j^*| \leq G$. Therefore, with $M_j(T) = (f_j^*)^{-1}(A_T)$, we have that $\mathcal{L}^1(M_j(T)) \leq \delta$ as well.

Returning to $K(T)$ and $K_j(T)$, we have that

$$K_j(T) \leq \int \chi_{\{f_j^* \in M_j(T)\}} \Psi(|\nabla f_j^*|) d\mathcal{L}^n$$

and similarly for $K(T)$. By the equi-absolute continuity result of Theorem 3.3, all of these numbers can simultaneously be made to be $\leq \frac{\gamma}{C}$ merely by choosing δ to be small enough. This, however, is accomplished by making T large enough. ■

Now we are in a position to prove the main theorem of this section.

7.5. Theorem ($\mathbf{W}^{1,1}$ convergence implies convergence for convex integrands and $\mathcal{E}_1$ functions). Suppose that the functions $f, f_1, f_2, f_3, \dots : \mathbf{R}^n \rightarrow \mathbf{R}^+$ belong to $\mathcal{E}_1$ and that $f_j \rightarrow f$ in measure as $j \rightarrow \infty$. Suppose also that $\Psi : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ is a suitable integrand and that $\Psi(|\nabla f|)$ and the $\Psi(|\nabla f_j|)$'s are $\mathcal{L}^n$ summable. Suppose additionally that

$$(a) \int \Psi(|\nabla f_j - \nabla f|) d\mathcal{L}^n \rightarrow 0 \text{ as } j \rightarrow \infty.$$

Then

$$(1) \quad \int |\Psi(|\nabla f_j|) - \Psi(|\nabla f|)| d\mathcal{L}^n \rightarrow 0 \text{ as } j \rightarrow \infty$$

and

$$(2) \text{ for every } 0 < Y < Z < \infty \text{ with } I = (Y, Z), \text{ we have}$$

$$\sup_j \mathcal{L}^n\{x : f_j^I(x) \neq 0\} < \infty \quad \text{and} \quad \mathcal{L}^n\{x : f^I(x) \neq 0\} < \infty$$

and

$$f_j^I \rightarrow f^I \text{ in } \mathbf{W}^{1,1}(\mathbf{R}^n) \text{ as } j \rightarrow \infty.$$

Now assume finally

$$(b) \text{ for every } Y \text{ and } Z \text{ with } 0 < Y < Z < \infty \text{ and with } I = (Y, Z), \text{ we have}$$

$$\int |\nabla f_j^{*I} - \nabla f^{*I}| d\mathcal{L}^n \rightarrow 0 \text{ as } j \rightarrow \infty.$$

Then

$$(3) \quad \int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \rightarrow 0 \text{ as } j \rightarrow \infty.$$

$$(4) \quad \int |\Psi(|\nabla f_j^*|) - \Psi(|\nabla f^*|)| d\mathcal{L}^n \rightarrow 0 \text{ as } j \rightarrow \infty.$$

Proof. It is well known that the assumed convergence of the f_j 's in measure implies the pointwise almost everywhere convergence of subsequences required by the hypotheses of Theorem 7.3. By Theorem 2.5, it also implies the convergence of the f_j^* 's to f^* in measure. (1) of Theorem 7.3 then implies (1) above and also implies that (3) above implies (4).

We now consider (2). We note (since $f_{(j)} \in \mathcal{C}_1$) that the set $A \equiv \{x : f(x) > \frac{Y}{4}\}$ has finite $\mathcal{L}^n$ measure. Let $B_j \equiv \{x : |f_j(x) - f(x)| > \frac{Y}{4}\}$. Since $f_j \rightarrow f$ in measure, $\mathcal{L}^n(B_j) < 1$ for all large j . Therefore, with $D_j \equiv \{x : f_j(x) > Y$ and $f(x) \leq Y\}$, we have

$$\mathcal{L}^n(D_j) \leq \mathcal{L}^n(A) + \mathcal{L}^n(B_j) \leq \mathcal{L}^n(A) + 1.$$

Thus,

$$\begin{aligned} \mathcal{L}^n\{x : |\nabla f_j^I(x) - \nabla f^I(x)| \neq 0\} &\leq \mathcal{L}^n\{x : f_j^I(x) \neq 0 \text{ or } f^I(x) \neq 0\} \\ &\leq \mathcal{L}^n(A) + \mathcal{L}^n(D_j) \\ &\leq 2\mathcal{L}^n(A) + 1. \end{aligned}$$

This is our initial assertion in (2).

We now prove that $|\nabla f_j^I - \nabla f^I| \rightarrow 0$ in $\mathbf{L}^1(\mathbf{R}^n)$. From the convexity and positivity properties of Ψ in Definition 2.1, we infer for each $\varepsilon > 0$ the existence of $C_\varepsilon < \infty$ such that $t \leq C_\varepsilon \Psi(t) + \varepsilon$ for each $t > 0$. Hence,

$$\int |\nabla f_j^I - \nabla f^I| d\mathcal{L}^n \leq C_\varepsilon \int \Psi(|\nabla f_j^I - \nabla f^I|) d\mathcal{L}^n + \varepsilon(2\mathcal{L}^n(A) + 1).$$

We first let $j \rightarrow \infty$, noting that the assumed convergence in measure together with hypothesis (a) establishes the hypotheses of Theorem 7.3. Consequently, by (2) of Theorem 7.3, we have that the integral on the right side of the equation above converges to 0 as $j \rightarrow \infty$. Now letting $\varepsilon \rightarrow 0$, we see that $|\nabla f_j^I - \nabla f^I| \rightarrow 0$ in $\mathbf{L}^1(\mathbf{R}^n)$.

To prove that $f_j^I \rightarrow f^I$ in $\mathbf{L}^1(\mathbf{R}^n)$, we first note as in the proof of Theorem 2.5 that $|f_j^I(x) - f^I(x)| \leq |f_j(x) - f(x)|$ for every x , and thus $f_j \rightarrow f$ in measure as well. For any $\varepsilon > 0$, we have for all large enough j that $|f_j^I(x) - f^I(x)| < \varepsilon$ except on a set of measure less than ε . Thus, for large j ,

$$\int |f_j^I - f^I| d\mathcal{L}^n \leq \varepsilon(2\mathcal{L}^n(A) + 1) + \varepsilon(Z - Y),$$

which establishes the $\mathbf{L}^1(\mathbf{R}^n)$ convergence.

We now consider (3). The falsity of (3) and the assumed convergence in measure would imply the existence of a further subsequence of the j 's (still denoted j) such that $f_j \rightarrow f$ and $f_j^* \rightarrow f^*$ pointwise almost everywhere as $j \rightarrow \infty$ while for each j ,

$$\int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \geq 7C^3\alpha > 0;$$

here C is the constant associated with Ψ as in Definition 2.1.

For $0 < Y < Z < \infty$, we set $H = (0, Y)$, $I = (Y, Z)$, $J = (Z, \infty)$, and check (using the fact that $\nabla f = 0$ almost everywhere on the inverse image of a point) that

$$\int \Psi(|\nabla f|) d\mathcal{L}^n = \int \Psi(|\nabla f^H|) d\mathcal{L}^n + \int \Psi(|\nabla f^I|) d\mathcal{L}^n + \int \Psi(|\nabla f^J|) d\mathcal{L}^n$$

and similarly with f replaced by the f_j 's and also

$$\begin{aligned} \int \Psi(|\nabla f^*|) d\mathcal{L}^n &= \int \Psi(|\nabla f^{*H}|) d\mathcal{L}^n + \int \Psi(|\nabla f^{*I}|) d\mathcal{L}^n \\ &\quad + \int \Psi(|\nabla f^{*J}|) d\mathcal{L}^n. \end{aligned}$$

Note here that $f^{*H} = f^{H*}$, etc.

We then use Definition 2.1 to estimate

$$\begin{aligned} \Psi(|\nabla f_j^* - \nabla f^*|) &= \Psi(|\nabla f_j^{*H} + \nabla f_j^{*I} + \nabla f_j^{*J} - \nabla f^{*H} - \nabla f^{*I} - \nabla f^{*J}|) \\ &\leq \Psi(5|\nabla f_j^{*I} - \nabla f^{*I}|) + \Psi(5|\nabla f_j^{*H}|) + \Psi(5|\nabla f_j^{*J}|) \\ &\quad + \Psi(5|\nabla f^{*H}|) + \Psi(5|\nabla f^{*J}|) \\ &\leq C^3[\Psi(|\nabla f_j^{*I} - \nabla f^{*I}|) + \Psi(|\nabla f_j^{*H}|) + \Psi(|\nabla f_j^{*J}|) \\ &\quad + \Psi(|\nabla f^{*H}|) + \Psi(|\nabla f^{*J}|)] \end{aligned}$$

so that (by Theorem 2.7)

$$\begin{aligned} 7\alpha &\leq \frac{1}{C^3} \int \Psi(|\nabla f_j^* - \nabla f^*|) d\mathcal{L}^n \\ &\leq \int \Psi(|\nabla f_j^{*I} - \nabla f^{*I}|) d\mathcal{L}^n + \int \Psi(|\nabla f_j^{*H}|) d\mathcal{L}^n \\ &\quad + \int \Psi(|\nabla f_j^{*J}|) d\mathcal{L}^n + \int \Psi(|\nabla f^{*H}|) d\mathcal{L}^n + \int \Psi(|\nabla f^{*J}|) d\mathcal{L}^n. \end{aligned}$$

Clearly, for every sufficiently small $Y > 0$ and sufficiently large $Z < \infty$, we have

$$\int \Psi(|\nabla f^H|) d\mathcal{L}^n < \alpha, \quad \int \Psi(|\nabla f^J|) d\mathcal{L}^n < \alpha.$$

For each such Y, Z , we use (3) of Theorem 7.3 to conclude for all large j ,

$$\int \Psi(|\nabla f_j^{*H}|) d\mathcal{L}^n < 2\alpha, \quad \int \Psi(|\nabla f_j^{*J}|) d\mathcal{L}^n < 2\alpha$$

so that for all large j ,

$$(*) \quad \alpha \leq \int \Psi(|\nabla f_j^{*I} - \nabla f^{*I}|) d\mathcal{L}^n.$$

The functions f_j^{*I} and f_j^{*J} are in $\mathcal{E}_2$ because f and f_j are in $\mathcal{E}_1$ (and hence by Theorem 2.6, f^* and f_j^* are in $\mathcal{E}_1$) and $\mathcal{L}^n\{x: f^*(x) > Y\} < \infty$. By hypothesis (a) and Theorem 7.3, $\Psi(|\nabla f_j^I - \nabla f^I|) \rightarrow 0$ in $L^1(\mathbf{R}^n)$ as $j \rightarrow \infty$.

Since $|\nabla f_j^I - \nabla f^I| \rightarrow 0$ in $L^1(\mathbf{R}^n)$ (by (2)), we can apply Theorem 7.4 to the functions f_j^{*I} and f_j^{*J} , and we can use hypothesis (b) to conclude that $\Psi(|\nabla f_j^{*I} - \nabla f^{*I}|) \rightarrow 0$ in $L^1(\mathbf{R}^n)$, which contradicts equation (*) and thereby establishes (3). ■

§8. ARC LENGTH CONVERGENCE AND $W^{1,1}$ CONVERGENCE

The graph of an absolutely continuous function $y = f(x)$ defined over, say, a bounded interval is a curve in the plane whose arc length can be computed by the usual integral formula. If $g = f^{-1}$ exists and is also absolutely continuous, then the arc length of its graph must equal the arc length of the graph of f . This implies equality for two different first derivative integrals, one involving f' and the other involving g' . One of the main goals of this part is to show that this idea can be extended to apply to somewhat more general pairs of functions. In the following sections, we will define an “arc length integral” of a general monotone function. Such monotone functions have generalized inverses whose arc length integrals are the same.

Our first problem is to define the “arc length integral” (or arc length, for short) of a monotone function $f: \mathbf{R} \rightarrow \mathbf{R}^+$ over an interval $[a, b]$. As indicated above, normally this would be the elementary calculus integral $\int_{[a,b]} \sqrt{1 + (f')^2} d\mathcal{L}^1$. One difficulty with this definition is that it does not give the answer we need for monotone functions with discontinuities. Also there are ambiguities if f has jumps at either of the two endpoints a or b . A resolution of these difficulties suitable for our purposes has led us to the following definition.

8.1. Definition (Arc length over an interval). Let $f: \mathbf{R} \rightarrow \mathbf{R}^+$ be monotonically nondecreasing (or, alternatively, nonincreasing). Then f has a distributional derivative $f' = \mathcal{L}^1 \wedge u + \nu$; here u is the absolutely continuous part of f' while ν is the singular part. It is well known that u coincides with the classical derivative of f almost everywhere. We define the *arc length of f over a bounded interval $[a, b]$* to be the number

$$\int_{[a,b]} \sqrt{1 + u^2} d\mathcal{L}^1 + \int_{[a,b]} d\nu.$$

In case f is a nonincreasing function, we write $f' = \mathcal{L}^1 \wedge u - \nu$ and use the same formula. Note that the second integral in this definition is not determined by the values of f in $[a, b]$ alone; different extensions outside $[a, b]$ can give different values for $\nu\{a\}$, $\nu\{b\}$.

Our first goal is to deduce certain properties of arc length. We begin with an elementary proposition which will be useful in several contexts.

8.2. Proposition (Upper semicontinuity can imply continuity). Suppose that μ measures X and $h, h_1, h_2, h_3, \dots: X \rightarrow \mathbf{R}^+$ are μ summable functions such that

- (a) $\liminf_{j \rightarrow \infty} \int h_j d\mu \leq \int h d\mu$,
- (b) $\liminf_{j \rightarrow \infty} h_j(x) \geq h(x)$ for μ almost every x .

Then there is a subsequence $j(1), j(2), j(3), \dots$ of $1, 2, 3, \dots$ such that

$$\lim_{k \rightarrow \infty} h_{j(k)}(x) = h(x) \quad \text{for } \mu \text{ almost every } x.$$

Proof. We shall first show that the function $\varphi(x) = \liminf_{j \rightarrow \infty} h_j(x)$ coincides with $h(x)$ for μ almost every x . To do this, we set

$$g_j(x) = \min\{\varphi(x), h_j(x)\} \leq h_j(x)$$

so that $\lim_{j \rightarrow \infty} g_j(x) = \varphi(x)$ for every x . We use Fatou's lemma to check that

$$\begin{aligned} \int h \, d\mu &\geq \liminf_{j \rightarrow \infty} \int h_j \, d\mu \geq \liminf_{j \rightarrow \infty} \int g_j \, d\mu \\ &\geq \int \lim_{j \rightarrow \infty} g_j \, d\mu = \int \varphi \, d\mu \end{aligned}$$

which then implies that

$$\int (\varphi - h) \, d\mu \leq 0.$$

Since (b) implies $\varphi - h \geq 0$, we infer $\varphi(x) = h(x)$ for μ almost every x .

We write $(h_j - h)_+ = \max\{(h_j - h), 0\}$ and $(h_j - h)_- = \max\{-(h_j - h), 0\}$ so that $(h_j - h) = (h_j - h)_+ - (h_j - h)_-$. Assumption (b) above thus implies $\lim_{j \rightarrow \infty} (h_j - h)_-(x) = 0$ for μ almost every x . Since $0 \leq (h_j - h)_- \leq h$ and h is μ summable, Lebesgue's dominated convergence theorem implies

$$\lim_{j \rightarrow \infty} \int (h_j - h)_- \, d\mu = 0.$$

Since

$$\int (h_j - h) \, d\mu + \int (h_j - h)_- \, d\mu = \int (h_j - h)_+ \, d\mu$$

and the middle integral converges to 0 as $j \rightarrow \infty$, we infer from assumption (a) that

$$\liminf_{j \rightarrow \infty} \int (h_j - h)_+ \, d\mu \leq 0.$$

Since $(h_j - h)_+ \geq 0$, there is a subsequence such that the functions $(h_j - h)_+$ converge to zero in the L^1 norm. This implies the existence of a further subsequence $j(1), j(2), j(3), \dots$ of $1, 2, 3, \dots$ such that

$$\lim_{k \rightarrow \infty} (h_{j(k)}(x) - h(x))_+ = 0$$

for μ almost every x . Since also $\lim_{k \rightarrow \infty} (h_{j(k)}(x) - h(x))_- = 0$ for μ almost every x , we therefore infer $\lim_{k \rightarrow \infty} h_{j(k)}(x) = h(x)$ for μ almost every x as required. ■

8.3. Theorem (Arc length integrals and convergence of functions). Suppose $w_0, w_1, w_2, \dots$ are nondecreasing functions mapping $\mathbf{R}$ to $\mathbf{R}^+$ and have associated distribution first derivatives given by Radon measures $\mathcal{L}^1 \wedge u_j + \nu_j$ for each $j = 0, 1, 2, \dots$; here ν_j is singular with respect to $\mathcal{L}^1$ for each j . In case

$$\liminf_{j \rightarrow \infty} u_j(t) \geq u_0(t)$$

for almost $\mathcal{L}^1$ every t and

$$|w_j - w_0| \rightarrow 0 \quad \text{in } L^1_{\text{loc}}(\mathbf{R}) \text{ as } j \rightarrow \infty,$$

then

$$(1) \quad \lim_{j \rightarrow \infty} \left(\int \varphi \sqrt{1 + u_j^2} d\mathcal{L}^1 + \int \varphi d\nu_j \right) = \int \varphi \sqrt{1 + u_0^2} d\mathcal{L}^1 + \int \varphi d\nu_0$$

for each continuous function $\varphi : \mathbf{R} \rightarrow \mathbf{R}$ having compact support.

Proof. Step 1. One infers from the monotonicity of the w_j 's together with the $L^1_{\text{loc}}(\mathbf{R})$ convergence of the w_j 's to w_0 as $j \rightarrow \infty$ that $w_j(t) \rightarrow w_0(t)$ as $j \rightarrow \infty$ at each point t of continuity of w_0 ; almost every t is a point of differentiability of w_0 , hence, in particular, a point of continuity.

For each smooth $\psi : \mathbf{R} \rightarrow \mathbf{R}$ with compact support, it follows from the definition of distribution derivative that

$$\int \psi u_j d\mathcal{L}^1 + \int \psi d\nu_j = - \int \psi' w_j d\mathcal{L}^1.$$

The $L^1_{\text{loc}}(\mathbf{R})$ convergence of the w_j 's to w_0 as $j \rightarrow \infty$ implies for each such ψ

$$(a) \quad \lim_{j \rightarrow \infty} \left(\int \psi u_j d\mathcal{L}^1 + \int \psi d\nu_j \right) = \int \psi u_0 d\mathcal{L}^1 + \int \psi d\nu_0.$$

Since the measures $\mathcal{L}^1 \wedge u_j + \nu_j$ are Radon measures which are locally uniformly bounded (since the w_j 's are monotone and converge), equation (a) holds if we replace ψ by a general continuous function φ with compact support. We conclude then, by subtracting (a) from (1), that it suffices to prove

$$(b) \quad \lim_{j \rightarrow \infty} \int \varphi G(u_j) d\mathcal{L}^1 = \int \varphi G(u_0) d\mathcal{L}^1$$

with $G(u) \equiv \sqrt{1 + u^2} - u \leq 1$. Clearly, it suffices to prove (b) for a subsequence.

Step 2. Let I be a bounded open interval containing the support of φ . Since ν_0 is singular, there is a measurable set A such that $\mathcal{L}^1(A) = \mathcal{L}^1(I)$ and $\nu_0(A) = 0$. By the regularity of Lebesgue measure, for every $\varepsilon > 0$ there is a closed set $B_\varepsilon \subset A$ such that $\mathcal{L}^1(B_\varepsilon) \geq \mathcal{L}^1(I) - \varepsilon$. We claim that

$$(c) \quad \liminf_{j \rightarrow \infty} \int_{B_\varepsilon} u_j d\mathcal{L}^1 \leq \int_{B_\varepsilon} u_0 d\mathcal{L}^1.$$

Assuming (c) for the moment, our hypothesis, together with Proposition 8.2, implies the existence of a subsequence (also denoted by u_j) such that $u_j \rightarrow u$, $\mathcal{L}^1 \llcorner B_\varepsilon$ almost everywhere. Then by Lebesgue's dominated convergence theorem,

$$(d) \quad \lim_{j \rightarrow \infty} \int_{B_\varepsilon} \varphi G(u_j) d\mathcal{L}^1 = \int_{B_\varepsilon} \varphi G(u_0) d\mathcal{L}^1$$

for this subsequence. We now pick such subsequences for a countable set of ε 's converging to 0 (so that (d) holds for each of these ε 's). Since φ is fixed and bounded and $G(u) \leq 1$, (b) follows readily from the truth of (d) for each of these ε 's. This completes the proof.

Step 3. To prove (c), we take any continuous function $\varphi : \mathbf{R} \rightarrow [0, 1]$ with compact support such that $\varphi(t) = 1$ for $t \in B_\varepsilon$ and $0 \leq \varphi(t) < 1$ for $t \notin B_\varepsilon$. For each positive integer m , we use (a) with ψ replaced by φ^m to obtain

$$\begin{aligned} \liminf_{j \rightarrow \infty} \int_{B_\varepsilon} u_j d\mathcal{L}^1 &\leq \liminf_{j \rightarrow \infty} \int \varphi^m u_j d\mathcal{L}^1 \\ &\leq \int \varphi^m u_j d\mathcal{L}^1 + \int \varphi^m d\nu_j \\ &\leq \int \varphi^m u_0 d\mathcal{L}^1 + \int \varphi^m d\nu_0. \end{aligned}$$

As $m \rightarrow \infty$, the first integral on the right converges (by dominated convergence) to $\int_{B_\varepsilon} u_0 d\mathcal{L}^1$ while the second integral converges to zero (since $\nu_0(B_\varepsilon) = 0$ and ν_0 is bounded). ■

8.4. Generalized graphs and inverses of monotone functions and their arc length integrals. Suppose a and b are positive numbers and that f and g are functions which map $\mathbf{R}$ to $\mathbf{R}$ and satisfy the following four conditions:

- (a) $f(x) = a$ for $-\infty < x \leq 0$, $f(x) = 0$ for $b \leq x < \infty$, $g(y) = b$ for $-\infty < y \leq 0$, $g(y) = 0$ for $a \leq y < \infty$;
- (b) $f(w) \geq f(x)$ whenever $-\infty < w < x < \infty$;
- (c) f is lower semicontinuous in the open interval $(0, b)$;
- (d) $g(y) = \sup\{x : f(x) > y\}$ for each $0 < y < a$.

We regard the restriction of g to $[0, a]$ as the *generalized inverse function* of the restriction of f to $[0, b]$. Then

- (1) $g(y) \geq g(z)$ whenever $-\infty < y < z < \infty$.

Proof. The failure of this assertion would imply the existence of $\varepsilon > 0$ and $y < z$ with $g(y) = g(z) - 2\varepsilon$. The definition of $g(z)$ implies the existence of $x > g(z) - \varepsilon$ with $f(x) > z$. Hence, $f(x) > y$, which implies $g(y) \geq x$. This is not possible since it would imply $g(z) - 2\varepsilon = g(y) \geq x > g(z) - \varepsilon$, which is false. ■

Also,

- (2) g is lower semicontinuous in the open interval $(0, a)$.

Proof. The failure of this assertion would imply the existence of $y_j \downarrow y \in (0, a)$ as $j \rightarrow \infty$ with $\lim_{j \rightarrow \infty} g(y_j) = g(y) - 2\varepsilon$ for some $\varepsilon > 0$. From the definition of $g(y)$, there is an increasing sequence x_j such that $f(x_j) > y$ and $g(y) = \lim_{j \rightarrow \infty} x_j$. Therefore, one of these x_j 's (call it x) satisfies the inequality $x > g(y) - \varepsilon$ with $f(x) \geq y + \delta$ for some $\delta > 0$. Hence, from the definition of $g(y)$, we have $g(y_j) \geq x > g(y) - \varepsilon$ whenever $y_j - y < \delta$, contrary to our assumption. ■

Finally,

- (3) $f(x) = \sup\{y : g(y) > x\}$ for each $0 < x < b$.

Proof. Our assertion is that for each $0 < x < b$,

$$f(x) = \sup\{y : \sup\{w : f(w) > y\} > x\}.$$

If the left side should be less than the right side, there would be $y > f(x)$ such that $\sup\{w : f(w) > y\} > x$, i.e., there would be $w > x$ with $f(w) > y > f(x)$ which is at variance with assumption (b) above. If the left side should be greater than the right side, there would be $\varepsilon > 0$ such that $\sup\{w : f(w) > y\} \leq x$ whenever $y \geq f(x) - 2\varepsilon$. The lower semicontinuity of f at x would imply the existence of $\delta > 0$ such that $f(w) > f(x) - \varepsilon$ whenever $|x - w| \leq \delta$. Hence, $\sup\{w : f(w) > f(x) - \varepsilon\} \geq x + \delta$, contrary to our earlier inequality, a special case of which is $\sup\{w : f(w) > f(x) - \varepsilon\} \leq x$. ■

We let $\rho : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ interchange x and y coordinates, i.e., $\rho(x, y) = (y, x)$ for each x, y . Then

(4) *The set*

$$\Gamma = [0, b] \times [0, a] \cap \left\{ (x, y) : \liminf_{t \rightarrow x} f(t) \leq y \leq \limsup_{t \rightarrow x} f(t) \right\}$$

coincides with the image under ρ of the set

$$\Sigma = [0, a] \times [0, b] \cap \left\{ (y, x) : \liminf_{s \rightarrow y} g(s) \leq x \leq \limsup_{s \rightarrow y} g(s) \right\}.$$

Proof. In view of the symmetry of the properties of f and g , it is sufficient to show that the $\rho(\Sigma) \subset \Gamma$. Suppose then that $0 < y < a$ and

$$\liminf_{s \rightarrow y} g(s) \leq x \leq \limsup_{s \rightarrow y} g(s).$$

We wish to show $0 \leq x \leq b$ (obvious) and

$$\liminf_{t \rightarrow x} f(t) \leq y \leq \limsup_{t \rightarrow x} f(t).$$

To see this, suppose first that $y < \liminf_{t \rightarrow x} f(t)$. This would imply

$$\liminf_{s \rightarrow y} g(s) \equiv \liminf_{s \rightarrow y} (\sup\{w : f(w) > s\}) > x,$$

contrary to our assumptions. Alternatively, suppose that $y > \limsup_{t \rightarrow x} f(t)$. This, in turn, would imply

$$\limsup_{s \rightarrow y} (\sup\{w : f(w) > s\}) < x,$$

again contrary to our assumptions. ■

We now wish to construct simple nonincreasing functions f_k and g_k which map $\mathbf{R}$ to $\mathbf{R}$ for $k = 1, 2, 3, \dots$ and which approximate f and g . We will describe the construction of the f_k 's. The construction of the g_k 's is the same except that f is replaced by g and Γ is replaced by Σ . We now proceed with the construction of f_k . We assert

(5) *Whenever $0 \leq w < x \leq b$ and $0 \leq y < z \leq a$, at most one of the two points (w, y) and (x, z) can belong to Γ . Furthermore, for each $-a \leq z \leq b$, there is exactly one point (x, y) in Γ with $y = x - s$.*

Proof. The first claim follows readily from (b) above. This fact then implies that there can be at most one point $(x, x - s) \in \Gamma$. To see that there is such a point we set

$$x = \sup\{t : t \leq b \text{ and } f(t) \leq t - s\}$$

and check from the definitions that

$$\liminf_{t \rightarrow x} f(t) \leq x - s \leq \limsup_{t \rightarrow x} f(t)$$

so that $(x, y) = (x, x - s) \in \Gamma$. ■

We let $\theta : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ denote counterclockwise rotation of the xy -plane by 45° and set $\Gamma^* = \theta(\Gamma)$. Then

(6) Γ^* is the graph of a function $h : [-a/\sqrt{2}, b/\sqrt{2}] \rightarrow [0, (a+b)/\sqrt{2}]$ which satisfies a Lipschitz condition with constant 1.

Proof. Since each of the lines $y = x - z$ of (5) above is rotated by θ to become vertical, the existence of h having graph Γ^* is clear. To check that $\text{Lip}(h) \leq 1$, we suppose otherwise and infer the existence of $(u, h(u)), (u + \Delta u, h(u) + \Delta h) \in \Gamma^*$ with $|\Delta h| > |\Delta u| > 0$. Then we set $(w, y) = \theta^{-1}(u, h(u))$, $(x, z) = \theta^{-1}(u + \Delta u, h(u) + \Delta h) \in \Gamma$, and compute $x - w = \Delta u + \Delta h$, $z - y = -\Delta u + \Delta h$. This implies that $x - w$ and $z - y$ have the same sign, which is incompatible with (5) above. ■

We let $a/\sqrt{2} = s_0 < s_1 < \dots < s_{2^k} = b/\sqrt{2}$ be equally spaced points partitioning the interval $[-a/\sqrt{2}, b/\sqrt{2}]$ into 2^k equal subintervals, each of length $(a+b)/2^k\sqrt{2}$. We set $p_j^* = (s_j, h(s_j)) \in \Gamma^*$ for $j = 0, \dots, 2^k$ and let Γ_k^* denote the polygonal figure which is the union of the line segments connecting p_{i-1}^* to p_i^* for $i = 1, \dots, 2^k$. Since $\text{Lip}(h) \leq 1$, each of these segments has length not exceeding $(a+b)/2^k$. The total length L_k of Γ_k^* does not exceed $(a+b)$ and is an approximation to the arc length L of Γ^* . Indeed, as is well known, the numbers L_k converge to L as $k \rightarrow \infty$ (for any reasonable definition of arc length). We further set $\gamma_k = \theta^{-1}(\Gamma_k^*)$, $(x_i, y_i) = p_i = \theta^{-1}(p_i^*) \in \Gamma_k$ for $i = 0, 1, \dots, 2^k$, and $K_k = \{x_0, x_1, \dots, x_{2^k}\}$ (there may be repetitions among the x_i 's). It is then straightforward to check

(7) For each $i = 0, 1, \dots, 2^k$, we have $\liminf_{t \rightarrow x_i} f(t) \leq y_i \leq \limsup_{t \rightarrow x_i} f(t)$ (since $(x_i, y_i) \in \Gamma$). Additionally, for each $i = 1, \dots, 2^k$, we have $x_{i-1} \leq x_i$ with $x_i - x_{i-1} \leq (a+b)/2^k$ and $y_{i-1} \geq y_i$ with $y_{i-1} - y_i \leq (a+b)/2^k$. Finally, for each $x \in (0, b) \sim K_k$, there is a unique $i \in \{1, \dots, 2^k\}$ such that $x_{i-1} < x < x_i$.

We define our function $f_k : \mathbf{R} \rightarrow \mathbf{R}$ by requiring

- (e) $f_k(w) = f(w)$ if $w \notin [0, b]$, and $(x, f_k(x)) \in \Gamma_k$ if $x \in [0, b]$,
- (f) f_k is lower semicontinuous in the open interval $(0, b)$.

One infers from (7) that

(8) f_k is well defined and is nondecreasing. Furthermore, if $i \in \{1, \dots, 2^k\}$ and $x_{i-1} < x < x_i$, then

$$f_k(x) = y_{i-1} + \left(\frac{y_i - y_{i-1}}{x_i - x_{i-1}} \right) (x - x_{i-1}).$$

We now denote by $g_k : \mathbf{R} \rightarrow \mathbf{R}$ the function analogous to f_k which one obtains by replacing f by g and Γ by Σ in the construction of f . Since the functions f, f_k, g, g_k are all nondecreasing, their classical first derivatives f', f'_k, g', g'_k exist $\mathcal{L}^1$ almost everywhere, and also their distribution first derivatives can then be written $\mathcal{L}^1 \wedge f' - \mu, \mathcal{L}^1 \wedge f'_k - \mu_k, \mathcal{L}^1 \wedge g' - \nu, \mathcal{L}^1 \wedge g'_k - \nu_k$, where μ, μ_k, ν, ν_k are Radon measures on $\mathbf{R}$ which are singular with respect to $\mathcal{L}^1$ (μ and μ_k have supports in $[0, b]$, and ν and ν_k have supports in $[0, a]$). Since arc length is invariant under rotations and Γ_k is a finite polyhedron (which may contain vertical segments), it is straightforward to check from the definition of distribution derivative that

$$(9) \quad L_k = \int_{[0, b]} \sqrt{1 + (f'_k)^2} d\mathcal{L}^1 + \int d\mu_k.$$

We let Σ_k denote the polygonal figure analogous to Γ_k obtained in the construction of g_k . It is then straightforward to check

$$(10) \quad \Sigma_k = \rho(\Gamma_k).$$

Since arc length is invariant under the reflection ρ , we similarly can write

$$(11) \quad L_k = \int_{[0, a]} \sqrt{1 + (g'_k)^2} d\mathcal{L}^1 + \int d\nu_k \quad (\text{same } L_k).$$

Clearly, $K_k \subset K_{k+1}$ for each k , and $K = \bigcup_k K_k \subset [0, b]$ is countable. We assert

(12) *If the classical derivative f' exists at $x \in \mathbf{R} \sim K$, then $\lim_{k \rightarrow \infty} f_k(x) = f(x)$ and $\lim_{k \rightarrow \infty} f'_k(x) = f'(x)$. Similarly, if the classical derivative g' exists at $y \in \mathbf{R} \sim K$, then $\lim_{k \rightarrow \infty} g_k(x) = g'(x)$ and $\lim_{k \rightarrow \infty} g'_k(x) = g'(x)$.*

Proof. In view of the symmetry of the properties of f and g and of our constructions, it is clearly adequate to establish our assertion for f alone. Since $f(x) = f_k(x)$ for $x \notin (0, b)$, it is sufficient to consider $x \in [0, b] \sim K$. We then infer from the definition of strong derivative the existence of a non-decreasing function $\varepsilon : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ with $\lim_{r \rightarrow 0+} \varepsilon(r) = 0$ such that for each $x \neq u \in \mathbf{R}$

$$\begin{aligned} f'(x) - \varepsilon(|u - x|) &< \liminf_{t \rightarrow u} \frac{f(t) - f(x)}{u - x} \\ &\leq \limsup_{t \rightarrow u} \frac{f(t) - f(x)}{u - x} \\ &< f'(x) + \varepsilon(|u - x|). \end{aligned}$$

For fixed k and $x \in (0, b) \sim K$, we let $x_{i-1} < x < x_i$ be as in (7) and use (7) and (8) to estimate

$$\begin{aligned}
 f'_k(x) &= \frac{y_i - y_{i-1}}{x_i - x_{i-1}} \\
 &= \frac{y_i - f(x) + f(x) - y_{i-1}}{x_i - x_{i-1}} \\
 &= \frac{[f'(x) \pm \varepsilon(|x_i - x|)](x_i - x) - [f'(x) \pm \varepsilon(|x_{i-1} - x|)](x_{i-1} - x)}{x_i - x_{i-1}} \\
 &= f'(x) \pm 2\varepsilon(|x_i - x_{i-1}|) \\
 &= f'(x) \pm 2\varepsilon((a+b)/2^k)
 \end{aligned}$$

which implies the required convergence of derivatives. For the convergence of the functions, we use (7) and (8) to estimate

$$\begin{aligned}
 |f_k(x) - f(x)| &= \left| y_{i-1} + \left(\frac{y_i - y_{i-1}}{x_i - x_{i-1}} \right) (x - x_{i-1}) - f(x) \right| \\
 &\leq \max\{|y_i - f(x)|, |y_{i-1} - f(x)|\} \\
 &\leq \max \left\{ \limsup_{t \rightarrow x_i} |f(t) - f(x)|, \limsup_{t \rightarrow x_{i-1}} |f(t) - f(x)| \right\}.
 \end{aligned}$$

Since f is differentiable at x , it is also continuous at x . (7) implies $x_i - x_{i-1}$ is small if k is large. The convergence of the functions follows. ■

In summary of (9), (11), and (12), we assert

(13) *We have the following convergence as $k \rightarrow \infty$:*

$$f_k(x) \rightarrow f(x) \quad \text{and} \quad g_k(x) \rightarrow g(x) \quad \text{for } \mathcal{L}^1 \text{ almost every } x,$$

$$f'_k(x) \rightarrow f'(x) \quad \text{and} \quad g'_k(x) \rightarrow g'(x) \quad \text{for } \mathcal{L}^1 \text{ almost every } x,$$

$$\begin{aligned}
 L_k &= \left(\int_{[0, b]} \sqrt{1 + (f'_k)^2} d\mathcal{L}^1 + \int d\mu_k \right) \\
 &= \left(\int_{[0, a]} \sqrt{1 + (g'_k)^2} d\mathcal{L}^1 + \int d\nu_k \right) \rightarrow L.
 \end{aligned}$$

As our primary conclusion, we have the following.

8.5. Theorem (Equality of arc length integrals). *Let f be a monotone function satisfying conditions (a), (b), (c) of 8.4, and let g be its “inverse function over $[0, b]$ ” satisfying conditions (a), (d) of 8.4. Then the arc length integral of f over $[0, b]$ equals the arc length integral of g over $[0, a]$, i.e.,*

$$\int_{[0, b]} \sqrt{1 + (f')^2} d\mathcal{L}^1 + \int_{[0, b]} d\mu = \int_{[0, a]} \sqrt{1 + (g')^2} d\mathcal{L}^1 + \int_{[0, a]} d\nu;$$

here μ and ν are the singular parts of the distribution first derivatives of f and g .

Proof. We use (13) of 8.4 above and Theorem 8.3. (13) states that the approximating functions f_k and g_k have the same arc lengths, so what we have to prove is that the arc length of f_k , for example, converges to the arc length of f . Since all the functions are bounded and $[0, b]$ is a finite interval, the pointwise convergence in (13) implies the L^1_{loc} convergence of f_k to f . Here it is important to note that for every k , $f_k(x) = f(x)$ for all x outside the interval $[0, b]$. Now choose some fixed continuous function φ with compact support such that $\varphi(x) = 1$ for $0 \leq x \leq b$. Since the derivatives of f_k converge to those of f (by (13)), the hypotheses of Theorem 8.3 are satisfied. The conclusion of Theorem 8.3 tells us that the arc lengths of the f_k 's over $[0, b]$ converge to the corresponding arc length of f except for the contributions in the integrals of (1) of Theorem 8.3 which come from $x < 0$ and $x > b$. Note, however, that the derivatives of all our functions vanish in these exterior regions, and therefore the external contributions are merely the integral of φ over $x < 0$ and $x > b$ —which is independent of k . ■

The following theorem states that strong convergence of gradients is a consequence of convergence of integrals of a strictly convex integrand. Theorem 8.7 contains the application needed for our main goal, Theorem 4.1. (1) of Theorem 8.6 is due to Serrin (see [SJ1, SJ2] or [GM, Theorem 2.11]). The other parts we believe to be new, although analogous results in geometric measure theory have been known for some time.

8.6. Theorem (Convergence with strictly convex integrands implies $W^{1,1}$ convergence). *Let $f, f_1, f_2, \dots$ be real-valued functions in $L^1_{\text{loc}}(\mathbf{R}^n)$, and assume that $f_j - f \rightarrow 0$ strongly in $L^1(\mathbf{R}^n)$. (It is not assumed that the f_j 's and f are separately summable.) Assume also that these functions have distributional gradients which are functions in $L^1_{\text{loc}}(\mathbf{R}^n, \mathbf{R}^n)$. Let $\Psi: \mathbf{R}^n \rightarrow \mathbf{R}^+$ be convex, and suppose that $\Psi(\nabla f)$ and the $\Psi(\nabla f_j)$'s are all summable. Then*

$$(1) \quad \liminf_{j \rightarrow \infty} \int \Psi(\nabla f_j) d\mathcal{L}^n \geq \int \Psi(\nabla f) d\mathcal{L}^n.$$

Now assume additionally that there is equality in (1) and that Ψ is strictly convex (i.e., $\Psi(x) + \Psi(y) > 2\Psi((x+y)/2)$ for all $x \neq y$ in $\mathbf{R}^n$). It is not assumed that Ψ is uniformly convex. Then there is a subsequence $j(k)$ of $1, 2, 3, \dots$ such that

$$(2) \quad \nabla f_{j(k)}(x) \rightarrow \nabla f(x) \quad \text{as } k \rightarrow \infty$$

for $\mathcal{L}^n$ almost every $x \in \mathbf{R}^n$. Moreover, for the original sequence,

$$(3) \quad \Psi(\nabla f_j) \rightarrow \Psi(\nabla f) \quad \text{as } j \rightarrow \infty$$

strongly in $L^1(\mathbf{R}^n)$ norm. If we assume additionally that $\Psi(\xi) \rightarrow \infty$ as $|\xi| \rightarrow \infty$, then for every measurable set $\Omega \subset \mathbf{R}^n$ with $\mathcal{L}^n(\Omega) < \infty$,

$$(4) \quad \nabla f_j \rightarrow \nabla f \quad \text{as } j \rightarrow \infty$$

strongly in $L^1(\Omega, \mathbf{R}^n)$ norm.

Proof. Step 1. Conclusion (1) seems more obvious than (4), but it is the key fact. We begin by proving that the more striking conclusions (2), (3), and (4) follow from (1) together with the equality assumption and the strict convexity assumption. Define the function

$$K_j(x) = \Psi(\nabla f(x)) + \Psi(\nabla f_j(x)) - 2\Psi\left(\frac{1}{2}(\nabla f(x) + \nabla f_j(x))\right)$$

which, by assumption, is strictly positive unless $\nabla f(x) = \nabla f_j(x)$. Consider the sequence $g_1, g_2, \dots$ with $g_j(x) = \frac{1}{2}[f(x) + f_j(x)]$. Clearly, $g_j - f \rightarrow 0$ in $L^1(\mathbf{R}^n)$, and since $K_j \geq 0$, we have further that $\Psi(\nabla g_j) \in L^1(\mathbf{R}^n)$. By (1) applied to the g_j 's,

$$\liminf_{j \rightarrow \infty} \int \Psi\left(\frac{1}{2}(\nabla f_j + \nabla f)\right) d\mathcal{L}^n \geq \int \Psi(\nabla f) d\mathcal{L}^n,$$

but by the assumed equality in (1) for $f, f_1, f_2, \dots$, we then have

$$\limsup_{j \rightarrow \infty} \int K_j(x) d\mathcal{L}^n \leq 0.$$

Since $K_j \geq 0$, we conclude that $K_j \rightarrow 0$ strongly in $L^1(\mathbf{R}^n)$. Since it suffices to prove (2), (3), and (4) for some subsequence, we choose a subsequence (still denoted by j) such that $K_j(x) \rightarrow 0$ for $\mathcal{L}^n$ almost every x . More precisely, there is a measurable set $A \subset \mathbf{R}^n$ of full $\mathcal{L}^n$ measure such that $K_j(x) \rightarrow 0$ for all $x \in A$.

We claim that the strict convexity of Ψ implies (2) for all $x \in A$. Suppose this is not true for some $x \in A$. Then there is a further subsequence such that either (i) for all j , $\nabla f_j(x)$ lies in some compact set B which does not contain $\nabla f(x)$, or else (ii) $|\nabla f_j(x)| \rightarrow \infty$. In (i), we can pass to a further subsequence such that $\nabla f_j(x) \rightarrow \xi \neq \nabla f(x)$. Since Ψ is continuous, $K_j(x)$ converges to $\Psi(\nabla f(x)) + \Psi(\xi) - 2\Psi(\frac{1}{2}(\nabla f(x) + \xi)) > 0$, which is a contradiction. Case (ii) can be reduced to (i) by the following exercise with convex functions. Fix y and $z \in \mathbf{R}^n$, and consider the convex function $M: (0, \infty) \rightarrow \mathbf{R}^+$ given by $M(\lambda) = \Psi(y + \lambda z)$. Since M is convex, we have that $M(\lambda) - 2M(\frac{\lambda}{2})$ is monotone nondecreasing in λ . (To see this easily, just differentiate: $M'(\lambda) - M'(\frac{\lambda}{2}) \geq 0$ since the derivative is a nondecreasing function.) Using this fact with $y = \nabla f(x)$ and $z = \nabla f_j(x) - \nabla f(x)$, we have that for all $0 \leq \lambda < 1$,

$$\begin{aligned} K_j(x) &\geq \Psi(\nabla f(x)) + \Psi(\nabla f(x) + \lambda(\nabla f_j(x) - \nabla f(x))) \\ &\quad - 2\Psi(\nabla f(x) + \tfrac{\lambda}{2}(\nabla f_j(x) - \nabla f(x))). \end{aligned}$$

Now choose $\lambda = \lambda_j = |\nabla f_j(x) - \nabla f(x)|^{-1}$, and we are back to (a) (i.e., $\xi_j \equiv [\nabla f_j(x) - \nabla f(x)]/|\nabla f_j(x) - \nabla f(x)|$ satisfies $|\xi_j| = 1$, so there is a subsequence such that $\xi_j \rightarrow \xi$ with $|\xi| = 1$, and we have the contradiction that $K_j(x) > 0$ in the limit).

We have proved (2) for some subsequence (which we continue to denote by j). To prove (3), we use the fact that $\|a\| - \|a - b\| \leq \|b\|$ for all real a, b .

Thus, if $\Phi \equiv \Psi(\nabla f_j) - |\Psi(\nabla f_j) - \Psi(\nabla f)|$, we have $|\Phi| \leq \Psi(\nabla f) \in \mathbf{L}^1(\mathbf{R}^n)$. As $j \rightarrow \infty$, by Lebesgue's dominated convergence theorem and the continuity of Ψ ,

$$\int [\Psi(\nabla f_j) - |\Psi(\nabla f_j) - \Psi(\nabla f)|] d\mathcal{L}^n \rightarrow \int \Psi(\nabla f) d\mathcal{L}^n.$$

But $\int \Psi(\nabla f_j) d\mathcal{L}^n \rightarrow \int \Psi(\nabla f) d\mathcal{L}^n$ by assumption, so (3) is proved.

The argument just given for $\Psi(\nabla f_j)$ can also be used to prove (4) for ∇f_j provided we can prove that for some subsequence $|\nabla f|$ and $|\nabla f_j|$ are in $\mathbf{L}^1(\Omega)$ for all j and that

$$(5) \quad \|\nabla f_j\|_{\mathbf{L}^1(\Omega)} \rightarrow \|\nabla f\|_{\mathbf{L}^1(\Omega)} \quad \text{as } j \rightarrow \infty.$$

We shall prove this when $\mathcal{L}^n(\Omega) < \infty$ and $\Psi(\xi) \rightarrow \infty$ as $|\xi| \rightarrow \infty$. This latter condition guarantees the existence of positive numbers T and C such that $C\Psi(\xi) > |\xi|$ whenever $|\xi| > T$. Since $\Psi(\nabla f)$ and $\Psi(\nabla f_j)$ are in $\mathbf{L}^1(\mathbf{R}^n)$, this fact implies that $|\nabla f|$ and $|\nabla f_j|$ are in $\mathbf{L}^1(\Omega)$. By Fatou's lemma, $\liminf_{j \rightarrow \infty} \|\nabla f_j\|_{\mathbf{L}^1(\Omega)} \geq \|\nabla f\|_{\mathbf{L}^1(\Omega)}$.

If (5) fails, then there is a further subsequence (again denoted by j) and $\gamma > 0$ such that

$$\|\nabla f_j\|_{\mathbf{L}^1(\Omega)} \geq \|\nabla f\|_{\mathbf{L}^1(\Omega)} + 4\gamma\mathcal{L}^n(\Omega)$$

for all j . By Egoroff's theorem, for every $\varepsilon > 0$, there is a measurable set $A_\varepsilon \subset \Omega$ with $\mathcal{L}^n(A_\varepsilon) < \varepsilon$ and an integer J_ε such that $|\nabla f_j(x)| \leq |\nabla f(x)| + \gamma$ for all $x \in \Omega \setminus A_\varepsilon$ and all $j > J_\varepsilon$. We can also assume that $A_\varepsilon \subset A_\delta$ when $\varepsilon > \delta$. There is then an $\varepsilon > 0$ such that $T\mathcal{L}^n(A_\varepsilon) < \gamma\mathcal{L}^n(\Omega)$ and (since $\Psi(\nabla f) \in \mathbf{L}^1(\mathbf{R}^n)$) $C \int_{A_\varepsilon} \Psi(\nabla f) d\mathcal{L}^n < \gamma\mathcal{L}^n(\Omega)$. We fix this ε and henceforth omit ε as a subscript. We then have for $j > J$ that

$$\begin{aligned} \|\nabla f_j\|_{\mathbf{L}^1(\Omega)} &\leq \|\nabla f\|_{\mathbf{L}^1(\Omega)} + \gamma\mathcal{L}^n(\Omega) + \int_A \chi_{\{|\nabla f_j| \leq T\}} |\nabla f_j| d\mathcal{L}^n \\ &\quad + \int_A \chi_{\{|\nabla f_j| > T\}} |\nabla f_j| d\mathcal{L}^n. \end{aligned}$$

The first integral on the right side is bounded above by $T\mathcal{L}^n(A) < \gamma\mathcal{L}^n(\Omega)$. For the second integral, we have that

$$|\nabla f_j| \leq C\Psi(\nabla f_j) \leq C\Psi(\nabla f) + C|\Psi(\nabla f_j) - \Psi(\nabla f)|.$$

Therefore, for $j > J$,

$$2\gamma\mathcal{L}^n(\Omega) \leq C \int_A \Psi(\nabla f) d\mathcal{L}^n + C \int_A |\Psi(\nabla f_j) - \Psi(\nabla f)| d\mathcal{L}^n,$$

and this is a contradiction because the first integral is bounded by $\gamma\mathcal{L}^n(\Omega)$ and the second converges to zero as $j \rightarrow \infty$.

Step 2. We turn now to the proof of (1). Obviously, we can assume (after passing to a subsequence) that the $\Psi(\nabla f_j)$'s are uniformly bounded in $\mathbf{L}^1(\mathbf{R}^n)$,

for otherwise there is nothing to prove. We begin by remarking that it suffices to prove

$$(6) \quad \liminf_{j \rightarrow \infty} \int g \Psi(\nabla f_j) d\mathcal{L}^n \geq \int g \Psi(\nabla f) d\mathcal{L}^n$$

for every nonnegative function $g \in C_c^\infty(\mathbf{R}^n)$. To prove this assertion, we can take $G \in C_c^\infty(\mathbf{R}^n)$ with $G(x) = 1$ for $|x| \leq 1$, $G(x) = 0$ for $|x| \geq 2$, and $0 \leq G(x) \leq 1$ otherwise. For later purposes, we also assume G to be spherical. Set $g_k(x) = G(\frac{x}{k})$ for $k = 1, 2, 3, \dots$. If (1) fails, there is a subsequence (again denoted by j) and some $\gamma > 0$ such that

$$\lim_{j \rightarrow \infty} \int \Psi(\nabla f_j) d\mathcal{L}^n < \int \Psi(\nabla f) d\mathcal{L}^n - 2\gamma.$$

Since $\Psi(\nabla f)$ is summable, Lebesgue's dominated convergence theorem implies that

$$\int g_k \Psi(\nabla f) d\mathcal{L}^n \rightarrow \int \Psi(\nabla f) d\mathcal{L}^n \quad \text{as } k \rightarrow \infty.$$

For some k , therefore,

$$\int g_k \Psi(\nabla f) d\mathcal{L}^n > \int \Psi(\nabla f) d\mathcal{L}^n - \gamma.$$

On the other hand, using (6) for this g_k , we have

$$\begin{aligned} \lim_{j \rightarrow \infty} \int \Psi(\nabla f_j) d\mathcal{L}^n &\geq \liminf_{j \rightarrow \infty} \int g_k \Psi(\nabla f_j) d\mathcal{L}^n \\ &\geq \int \Psi(\nabla f) d\mathcal{L}^n - \gamma, \end{aligned}$$

which is a contradiction.

Let K be the support of g , and let $B \supset K$ be bounded and open. We claim next that the general case of (6) follows from the special case that f_j and f are in $C^\infty(B)$ with gradients that are uniformly bounded in $L^1(B)$. To see this, we define $h_k(x) = \sigma k^{n/2} G(xk)$ for $k = 1, 2, 3, \dots$, where the constant σ is chosen so that $\int h_k d\mathcal{L}^n = 1$ for all k . We also define the infinitely differentiable approximants $f^{(k)} = h_k * f$ and $f_j^{(k)} = h_k * f_j$. These are well defined since f_j and $f \in L^1_{\text{loc}}(\mathbf{R}^n)$, and we have $\nabla f^{(k)} = h_k * \nabla f$ and $\nabla f_j^{(k)} = h_k * \nabla f_j$. We also have the bound $|\nabla f_k^{(k)}| \leq |\nabla h_k| * f_j$ so that $\|\nabla f_j^{(k)}\|_{L^1(B)} \leq C_k \|f_j\|_{L^1(B)}$ and similarly for $\nabla f^{(k)}$. Since $\nabla f \in L^1(B)$, $\nabla f^{(k)}$ converges to ∇f strongly in $L^1(B)$ as $k \rightarrow \infty$, so there is a subsequence of the k 's (which we continue to denote by k) such that $\nabla f^{(k)}(x) \rightarrow \nabla f(x)$ for $\mathcal{L}^n$ almost every $x \in B$. Then also $\Psi(\nabla f^{(k)}(x)) \rightarrow \Psi(\nabla f(x))$ for $\mathcal{L}^n$ almost every $x \in K$ since Ψ is continuous.

If (6) is false for f_j and f (but true for each $f_j^{(k)}$ and $f^{(k)}$), we have for some $\gamma > 0$ that

$$\liminf_{j \rightarrow \infty} \int g \Psi(\nabla f_j) d\mathcal{L}^n < \int g \Psi(\nabla f) d\mathcal{L}^n - 2\gamma.$$

By Fatou's lemma, there is an infinite subset, M , of the integers such that when $k \in M$,

$$\int g\Psi(\nabla f) d\mathcal{L}^n < \int g\Psi(\nabla f^{(k)}) d\mathcal{L}^n + \gamma,$$

and thus

$$\liminf_{j \rightarrow \infty} \int g\Psi(\nabla f_j) d\mathcal{L}^n < \int g\Psi(\nabla f^{(k)}) d\mathcal{L}^n - \gamma.$$

On the other hand, Jensen's inequality tells us that

$$h_k * \Psi(\nabla f_j) \geq \Psi(h_k * \nabla f_j) = \Psi(\nabla f_j^{(k)}).$$

Using the simple inequality

$$\begin{aligned} \liminf_{j \rightarrow \infty} \int g\Psi(\nabla f_j) d\mathcal{L}^n &\geq \liminf_{j \rightarrow \infty} \int (g * h_k)\Psi(\nabla f_j) d\mathcal{L}^n \\ &\quad + \liminf_{j \rightarrow \infty} \int (g - g * h_k)\Psi(\nabla f_j) d\mathcal{L}^n, \end{aligned}$$

we have by Fubini's theorem that the first integral on the right equals

$$\int g h_k * \Psi(\nabla f_j) d\mathcal{L}^n \geq \int g\Psi(\nabla f_j^{(k)}) d\mathcal{L}^n.$$

But by (6) as applied to $f_j^{(k)}$ and $f^{(k)}$,

$$\liminf_{j \rightarrow \infty} \int g\Psi(\nabla f_j^{(k)}) d\mathcal{L}^n \geq \int g\Psi(\nabla f^{(k)}) d\mathcal{L}^n,$$

and therefore we have that

$$\liminf_{j \rightarrow \infty} \int (g - g * h_k)\Psi(\nabla f_j) d\mathcal{L}^n < -\gamma$$

for $k \in M$. This is a contradiction because this integral is bounded above by

$$\sup_x \{g(x) - (g * h_k)(x)\} \sup_j \int \Psi(\nabla f_j) d\mathcal{L}^n < \infty.$$

As $k \rightarrow \infty$ (in M), $g(x) - (g * h_k)(x) \rightarrow 0$ uniformly in x , and hence for some fixed large k this integral exceeds $-\gamma$.

We can now concentrate on proving (6) for f_j and f infinitely differentiable. For this purpose, we can assume that Ψ is also infinitely differentiable. We merely replace Ψ by $\Psi^{(k)} \equiv h_k * \Psi$ which, since h_k is spherically symmetric, is also convex. Moreover, the ranges of ∇f_j and ∇f are now assumed to lie in some fixed compact set L in $\mathbf{R}^n$, and it is easy to see that $\Psi^{(k)} \rightarrow \Psi$ uniformly on L . (In fact, if $G(x)$ is also chosen to be decreasing in $|x|$, then $\Psi^{(k)}(x)$ is monotone nonincreasing in k for each $x \in \mathbf{R}^n$.)

Since Ψ is convex and smooth,

$$\Psi(\xi) - \Psi(\eta) \geq (\xi - \eta) \bullet \nabla \Psi(\eta)$$

for all ξ , and $\eta \in \mathbf{R}^n$. Thus, it suffices to prove that

$$I_j \equiv \int g(\nabla f_j - \nabla f) \bullet (\nabla \Psi(\nabla f)) d\mathcal{L}^n$$

converges to zero as $j \rightarrow \infty$. Since all these functions are in $C^\infty(B)$ and g has compact support, we can integrate by parts.

$$I_j = - \int (f_j - f) \operatorname{div}(g(\nabla \Psi(\nabla f))) d\mathcal{L}^n,$$

and this clearly converges to zero since $f_j - f \rightarrow 0$ in $\mathbf{L}^1(\mathbf{R}^n)$ and $\operatorname{div}(\cdots)$ is bounded. ■

Theorem 8.6 was proved in order to help establish the following Theorem 8.7 which is one of our principal results about arc length.

8.7. Theorem (Arc length convergence implies $\mathbf{W}^{1,1}(\mathbf{R})$ convergence). Suppose $v, v_1, v_2, v_3, \dots$ are nondecreasing absolutely continuous functions mapping the interval $[0, b]$ to the interval $[0, a]$ with

$$\lim_{j \rightarrow \infty} \int_{[0, b]} |v_j - v| d\mathcal{L}^1 = 0$$

and

$$\lim_{j \rightarrow \infty} \int_{[0, b]} \sqrt{1 + (v'_j)^2} d\mathcal{L}^1 = \int_{[0, b]} \sqrt{1 + (v')^2} d\mathcal{L}^1.$$

Then whenever $0 \leq \alpha < \beta \leq 1$, the following hold:

$$\lim_{j \rightarrow \infty} \int_{[\alpha, \beta]} \sqrt{1 + (v'_j)^2} d\mathcal{L}^1 = \int_{[\alpha, \beta]} \sqrt{1 + (v')^2} d\mathcal{L}^1$$

and

$$\lim_{j \rightarrow \infty} \int_{[\alpha, \beta]} |v'_j - v'| d\mathcal{L}^1 = 0.$$

Proof. Clearly, $\Psi(x) = \sqrt{1 + x^2} - 1$ is strictly convex, and it is bounded by $|x|$. Since v'_j and v' are in $\mathbf{L}^1([0, b])$, we can easily extend v_j and v to functions $\tilde{v}_j$ and $\tilde{v}$ defined in all of $\mathbf{R}$ such that $\tilde{v}_j - \tilde{v} \rightarrow 0$ in $\mathbf{L}^1(\mathbf{R})$ and $\Psi(\tilde{v}'_j)$ and $\Psi(\tilde{v}')$ are in $\mathbf{L}^1(\mathbf{R})$ with $\Psi(\tilde{v}_j) \rightarrow \Psi(\tilde{v})$ outside $[0, b]$; e.g., $\tilde{v}(x) = v(x)$ for $0 \leq x \leq b$, $\tilde{v}(x) = v(b)(b + 1 - x)$ for $b \leq x \leq b + 1$, $\tilde{v}(x) = v(0)(1 + x)$ for $-1 \leq x \leq 0$, and $\tilde{v}(x) = 0$ otherwise, and similarly for $\tilde{v}_j(x)$. Then apply Theorem 8.6.

With the foregoing proposition about arc length, we are now ready for the main results of this section.

8.8. Proof of Theorem 4.1. Assume that $f, f^*, \mathcal{V} = \mathcal{V}_f, \mathcal{W}, \nu$ and $f_j, f_j^*, \mathcal{V}_j = \mathcal{V}_{f_j}, \mathcal{W}_j$ have the meaning in hypotheses (a)–(g) of Theorem 4.1.

Conclusion (1) of Theorem 4.1 follows directly from our assumptions (a)–(d) and (1) of Theorem 7.5. Likewise, (4) follows from (2) together with (1) of Theorem 7.5 (with the help of Theorem 2.7).

(3) of Theorem 4.1 implies (2) of Theorem 4.1. By (2) of Theorem 7.5, we know that $f_j^I \rightarrow f^I$ in $\mathbf{W}^{1,1}(\mathbf{R}^n)$ as $j \rightarrow \infty$ for every interval $I = (Y, Z)$ with $0 < Y < Z < \infty$. Since (3) of Theorem 7.5 corresponds to (2) of Theorem 4.1, it suffices to establish hypothesis (b) of Theorem 7.5, i.e., to prove that

$\nabla f_j^{I*} \rightarrow \nabla f^{I*}$ in $L^1(\mathbf{R}^n)$ for every $I = (Y, Z)$. We now fix such $I = (Y, Z)$. Two facts must be remembered.

(1) f_j^I and f^{I*} are bounded above by the constant $a = Z - Y$.

(2) f_j^{*I} and f^{*I} vanish for $|x| \geq$ some constant R .

This latter fact follows from the first part of (2) of Theorem 7.5. We denote by $\mathcal{V}^I, \mathcal{W}^I$, etc. the distribution functions (Definition 1.2.5) of f^I , etc. Note that $\mathcal{V}^I = \mathcal{V}_{f^I}$, etc. We note that the assumed convergence of the f_j 's to f in measure together with hypothesis (d) of Theorem 4.1 establish the hypotheses of Theorem 7.3. Consequently, by (2) of Theorem 7.3, we have that assumption (2) of Theorem 4.1 holds with the f_j 's replaced by f_j^I 's, etc. We then readily conclude that the hypotheses (a)–(g) of Theorem 4.1 are satisfied for the f_j^I 's etc. Furthermore, (3) of Theorem 4.1 holds for the $\mathcal{W}_{f_j^I}$'s since it was assumed to hold for the $\mathcal{W}_j$'s. Assuming the failure of the $W^{1,1}$ convergence of the f_j^{I*} 's to f^{I*} , we can pass to a subsequence (still denoted by j 's) so that $\|f_j^{I*} - f^{I*}\|_{W^{1,1}} > \alpha > 0$ for all j 's.

Now consider the monotone nonincreasing distribution functions $y \mapsto \mathcal{V}_j^I(y)$ and $y \mapsto \mathcal{V}^I(y)$ which are defined *a priori* for $y > 0$ and which are zero when $y > a$. These we extend to all real y by setting $\mathcal{V}_{(j)}^I(y) = \alpha(n)R^n \equiv b$ for $y \leq 0$. (Note that $\mathcal{V}$ is the “volume” function which equals the volume of the unit n ball $\alpha(n)$ times the radius R raised to the n th power.)

We wish to apply Theorem 8.3 and, to this end, we first have to verify that $\mathcal{V}_j^I \rightarrow \mathcal{V}^I$ in $L^1([0, a])$. We define the $\mathcal{L}^{n+1}$ measurable sets

$$\Omega_j \equiv \mathbf{R}^n \times \mathbf{R} \cap \{(x, y) : f^I(x) < y < f_j^I(x) \text{ or } f_j^I(x) < y < f^I(x)\}$$

and infer from Fubini's theorem that

$$(3) \quad \|\mathcal{V}_j^I - \mathcal{V}^I\|_{L^1} \leq \mathcal{L}^{n+1}(\Omega_j) = \|f_j^I - f^I\|_{L^1}.$$

The right side converges to zero (by (2) of Theorem 7.5). Our desideratum is verified (by passing to a further subsequence if necessary) since (3) of Theorem 4.1 holds for the $\mathcal{W}_{(j)}^I$'s. Therefore, (1) of Theorem 8.3 holds. By imitating the proof of Theorem 8.5 to take care of the φ 's in Theorem 8.3, we deduce that (1) of Theorem 8.3 implies the convergence of the arc lengths of the $\mathcal{V}_j^I$'s to that of $\mathcal{V}^I$ on $[0, a]$.

Next we consider the functions $\mathcal{U}_{(j)}^I : \mathbf{R} \rightarrow [0, a]$ which are the “inverses” of the $\mathcal{V}_j^I$'s and $\mathcal{V}^I$ according to the definition given in 8.4. All these take the value zero if the argument s satisfies $s \geq b$ and the value a if $s \leq 0$. The $\mathcal{U}_{(j)}^I$'s are simply related to the $f_{(j)}^{I*}$'s for $s \geq 0$ by

$$(4) \quad \mathcal{U}_{(j)}^I(\alpha(n)r^n) = f_{(j)}^{I*}(r)$$

with $s = \alpha(n)r^n$. (Note that the $f_{(j)}^I$'s are also lower semicontinuous according to Definition 1.3.) By Theorem 8.5, the convergence of the arc lengths of the $\mathcal{V}_j^I$'s implies the convergence of the arc length of $\mathcal{U}_j^I$ to that of $\mathcal{U}^I$. Likewise, by the same argument as in (3) above, the $L^1([0, a])$ convergence of the $\mathcal{V}_j^I$'s to $\mathcal{V}^I$ implies the $L^1([0, b])$ convergence of the $\mathcal{U}_j^I$'s to $\mathcal{U}^I$. We also note that the $\mathcal{U}_{(j)}^I$'s are absolutely continuous on $(0, \infty)$ since the $f_{(j)}^*$'s are; i.e., the $f_{(j)}^*$'s are in $\mathbf{W}^{1,1}(\mathbf{R}^n)$ by Theorem 2.6. Then by Theorem 8.7, we can conclude

$$(5) \quad \mathcal{U}_j^I \rightarrow \mathcal{U}^I \quad \text{in } \mathbf{W}^{1,1}(0, \infty).$$

To complete the proof, we have to show that $\nabla f_j^{I*} \rightarrow \nabla f^{I*}$ in $L^1(\mathbf{R}^n)$ and thereby establish a contradiction with our initial assumptions about the failure of convergence. The derivatives of the $f_{(j)}^{I*}$ are related to those of the $\mathcal{U}_{(j)}^I$ by

$$(6) \quad (f_{(j)}^{I*})'(r) = n\alpha(n)r^{n-1}(\mathcal{U}_{(j)}^I)'(\alpha(n)r^n).$$

Thus,

$$\begin{aligned} \int |\nabla f_j^{I*} - \nabla f^{I*}| d\mathcal{L}^n &= n\alpha(n) \int_{0 < r < \infty} |(f_{(j)}^{I*})' - (f^{I*})'| (r) r^{n-1} d\mathcal{L}^1 r \\ &= n\alpha(n) \int_{0 < s < b} |(\mathcal{U}_j^I)' - (\mathcal{U}^I)'|(s) \left[\frac{s}{\alpha(n)} \right]^{1-1/n} d\mathcal{L}^1 s \end{aligned}$$

from which the desired convergence is evident.

(2) of Theorem 4.1 implies (3) of Theorem 4.1. If (3) of Theorem 4.1 is not true, then there must be some fixed slab $I = (Y, Z)$ with $0 < Y < Z < \infty$ for which $\mathcal{W}_{(j)}^I$ fails to satisfy (3) of Theorem 4.1. This, in turn, would imply for every subsequence that

$$(7) \quad \liminf_{j \rightarrow \infty} |\mathcal{W}_j^I(y)| < |\mathcal{W}^I(y)|$$

for all y in some set A of positive $\mathcal{L}^1$ measure. Here A may depend on the subsequence, but $\mathcal{L}^n(A) > 0$ always.

As a consequence of our assumption (2) of Theorem 4.1 and the convergence of f_j^{I*} to f^{I*} in measure (Theorem 2.4), Theorem 7.2 states that $f_j^{I*} \rightarrow f^{I*}$ in $\mathbf{W}^{1,1}(\mathbf{R}^n)$. Hence there is a subsequence such that $f_j^{I*} \rightarrow f^{I*}$ and $\nabla f_j^{I*} \rightarrow \nabla f^{I*}$ pointwise almost everywhere. However, (7) above continues to hold for this new sequence on some set A of positive measure, since it holds for every subsequence.

As before, we set $f_{(j)}^{I*}(r) = 0$ for $r \geq b = \text{fixed constant}$ (see (2) of Theorem 7.5). By Theorem 8.3, the arc length of f_j^{I*} converges to the arc length of f^{I*} on $[0, b]$. This arc length convergence together with the pointwise convergence of f_j^{I*} to f^{I*} implies (by Theorem 8.5) that

$$(8) \quad \int \left[\sqrt{1 + (\mathcal{W}_j^I)^2} - \mathcal{W}_j^I \right] d\mathcal{L}^1 \rightarrow \int \left[\sqrt{1 + (\mathcal{W}^I)^2} - \mathcal{W}^I \right] d\mathcal{L}^1$$

as $j \rightarrow \infty$. (This follows from the convergence of $\int \mathcal{W}_j^{I'} d\mathcal{L}^1$ to $\int \mathcal{W}^{I'} d\mathcal{L}^1$ which, in turn, follows from the pointwise convergence and the definition of distribution derivatives. See Step 1 of the proof of Theorem 8.3.)

The integrand in (8) above is strictly convex, and hence by Theorem 8.6, $\mathcal{W}_j^{I'} \rightarrow \mathcal{W}^{I'}$ in $L^1(\mathbf{R})$. As is well known, this implies convergence in measure of $\mathcal{W}_j$ to $\mathcal{W}$ and this implies that for some subsequence $\mathcal{W}_j$ converges to $\mathcal{W}$ pointwise almost everywhere; this contradicts the assertion that (7) above holds for every subsequence. ■

8.9. Proof of Theorem 4.2. We shall show that co-area regularity implies (3) of Theorem 4.1 and hence (2) of Theorem 4.1, which is our desired conclusion. It clearly suffices to do this for a subsequence of the j 's.

Consider the distribution functions $\mathcal{V}_{(j)} = \mathcal{V}_{f_{(j)}}$, $\mathcal{F}_{(j)} = \mathcal{F}_{f_{(j)}}$, $\mathcal{G}_{(j)} = \mathcal{G}_{f_{(j)}}$ of $f_{(j)}$ with $\mathcal{V}_{(j)} = \mathcal{F}_{(j)} + \mathcal{G}_{(j)}$; here (j) denotes j or no j (see Definition 1.2.5). Our co-area regularity assumption is that the distribution derivative of $\mathcal{F}_f$ is a purely singular Radon measure. We let $\mathcal{W}_{(j)}$ denote that part of the distribution derivative of $\mathcal{V}_{(j)}$ which is absolutely continuous with respect to $\mathcal{L}^1$.

For various slabs $I = (Y, Z)$, consider the corresponding slab functions $f_{(j)}^I$, $\mathcal{V}_{(j)}^I$, $\mathcal{W}_{(j)}^I$, $\mathcal{F}_{(j)}^I$, $\mathcal{G}_{(j)}^I$; here $\mathcal{F}_{(j)}^I$ is the co-area distribution function associated with $f_{(j)}^I$, etc. According to (2) of Theorem 7.5, the f_j^I 's converge to f^I in $\mathbf{W}^{1,1}(\mathbf{R}^n)$. According to Theorem 3.5, there is a subsequence of the j 's (still denoted by j) such that $\limsup_{j \rightarrow \infty} \mathcal{F}_j^{I'}(y) \leq \mathcal{F}_j^{I'}(y)$ for almost every y ; here $\mathcal{F}_{(j)}^{I'}$ is the derivative of the co-area distribution function associated with the $f_{(j)}^I$. As we remarked after Definition 7.1, the function f^I is also co-area regular, and therefore the distribution derivative of its residual distribution function $\mathcal{G}^{I'}$ is a purely singular Radon measure. Since any possible absolutely continuous part of the distribution derivative of the $\mathcal{G}_j^{I'}$'s is necessarily negative, the conclusion of Theorem 3.5 can only be strengthened, i.e.,

$$(1) \quad \limsup_{j \rightarrow \infty} \mathcal{W}_j^{I'} \leq \mathcal{W}^{I'}.$$

Since (1) holds for each I , it holds without any I , i.e., for $\mathcal{W}_{(j)}(y)$. The reason is the obvious fact that for each $I = (Y, Z)$, $\mathcal{V}_{(j)}(y) = \mathcal{V}_{(j)}^I(y - Y)$ and $\mathcal{W}_{(j)}(y) = \mathcal{W}_{(j)}^I(y - Y)$ whenever $Y < y < Z$. Hence (1) implies the desired conclusion (3) of Theorem 4.1. ■

§9. SYMMETRIC REARRANGEMENT IS STRONGLY CONTINUOUS FOR FRACTIONAL DERIVATIVES

9.1. Fractional Sobolev spaces. The main purpose of this paper is to delineate the conditions under which strong convergence of ∇f_j to ∇f leads to strong convergence of ∇f_j^* to ∇f^* . It was shown that the latter is a consequence of the former if and only if the derivative of $\mathcal{G}_f(y) \equiv \int \chi_{\{f > y\}} \chi_{\{\nabla f = 0\}} d\mathcal{L}^n$, as

a measure, does not have a piece that is absolutely continuous with respect to Lebesgue measure $\mathcal{L}^1$. The existence of an absolutely continuous piece is a delicate phenomenon that can occur only in dimensions $n \geq 2$.

It turns out that for fractional derivatives this kind of problem does not arise; the rearrangements *always* converge. It is not our intention to discuss fractional derivatives in full generality, and we shall content ourselves here with the discussion of one class of spaces, namely, the fractional Sobolev space $\mathbf{W}^{\alpha,p}(\mathbf{R}^n)$ with $0 < \alpha < 1$ and $1 \leq p < \infty$.

Recall that $\mathbf{W}^{1,p}(\mathbf{R}^n)$ is the space of functions $f: \mathbf{R}^n \rightarrow \mathbf{C}$ with the property that $f \in \mathbf{L}^p(\mathbf{R}^n)$ and ∇f , as a distribution, is a $\mathbf{C}^n$ -valued function in $\mathbf{L}^p(\mathbf{R}^n, \mathbf{C}^n)$. When $p = 2$, an equivalent characterization, in terms of the Fourier transform $\hat{f}$, is that both $\hat{f}$ and $|\xi|\hat{f}(\xi)$ are in $\mathbf{L}^2(\mathbf{R}^n)$. To define the space $\mathbf{W}^{\alpha,2}(\mathbf{R}^n)$, we merely replace $|\xi|$ by $|\xi|^\alpha$. For $\mathbf{W}^{1,2}(\mathbf{R}^n)$, there are the relations

$$\|f\|_{\mathbf{L}^2} = \|\hat{f}\|_{\mathbf{L}^2} \quad \text{and} \quad \|\nabla f\|_{\mathbf{L}^2} = \| |\xi| \hat{f} \|_{\mathbf{L}^2}.$$

The analogous relation for $\mathbf{W}^{\alpha,2}(\mathbf{R}^n)$ with $0 < \alpha < 1$ is the following in which C_α is a suitable constant.

$$C_\alpha \| |\xi|^\alpha \hat{f} \|_{\mathbf{L}^2}^2 = I^{\alpha,2}(f) \equiv \iint |f(x) - f(y)|^2 |x - y|^{-n-2\alpha} d\mathcal{L}^n x d\mathcal{L}^n y.$$

The definition of the $\mathbf{W}^{\alpha,p}$ norm in the $p \neq 2$ case imitates this. It is given by

$$\|f\|_{\mathbf{W}^{\alpha,p}}^p = \|f\|_p^p + I^{\alpha,p}(f),$$

$$I^{\alpha,p}(f) = \iint |f(x) - f(y)|^p |x - y|^{-n-p\alpha} d\mathcal{L}^n x d\mathcal{L}^n y.$$

Since

$$\int_0^\infty \exp[-|x|^2 t] t^{(n+\alpha p)/2-1} d\mathcal{L}^1 t = \Gamma((n+\alpha p)/2) |x|^{-n-\alpha p},$$

we have (by Fubini's theorem) that

$$I^{\alpha,p}(f) = \int G(f, t, p) d\mu^{\alpha,p} t,$$

$$G(f, t, p) = \iint |f(x) - f(y)|^p \exp[-|x - y|^2 t] d\mathcal{L}^n x d\mathcal{L}^n y,$$

and where $\mu^{\alpha,p}$ is the measure on $[0, \infty]$ given by

$$\mu^{\alpha,p} = \mathcal{L}^1 \wedge \Gamma((n+\alpha p)/2)^{-1} t^{(n+\alpha p)/2-1}.$$

9.2. Theorem (Rearrangement is continuous in $\mathbf{W}^{\alpha,p}$). Let $0 < \alpha < 1$ and $1 \leq p < \infty$. As in 2.8, define $f^* = |f|^*$.

- (1) If $f \in \mathbf{W}^{\alpha,p}(\mathbf{R}^n)$, then f^* is also in $\mathbf{W}^{\alpha,p}(\mathbf{R}^n)$ and $\|f^*\|_{\mathbf{W}^{\alpha,p}} \leq \|f\|_{\mathbf{W}^{\alpha,p}}$.
- (2) If f_j is a sequence of functions in $\mathbf{W}^{\alpha,p}(\mathbf{R}^n)$ and if $f_j \rightarrow f$ strongly in $\mathbf{W}^{\alpha,p}(\mathbf{R}^n)$ norm, then $f_j^* \rightarrow f^*$ strongly in $\mathbf{W}^{\alpha,p}(\mathbf{R}^n)$ norm.

Proof. Trivially, $\|f^*\|_{L^p} = \|f\|_{L^p}$. Also, by Corollary 2.3 (a rearrangement inequality for convex integrands), we have that $\|f_j^* - f^*\|_{L^p} \leq \|f_j - f\|_{L^p}$. Thus, it suffices for part (1) to prove that $G(f_j^*, t, p) \leq G(f, t, p)$ for all $t \geq 0$. This, however, is a consequence of the same corollary since $z \rightarrow |z|^p$ is convex and $x \rightarrow \exp[-|x|^2 t]$ is symmetric decreasing.

For part (2), we claim that it suffices to prove two things:

- (a) $\lim_{j \rightarrow \infty} G(f_j^* - f^*, t, p) = 0$ for all $t > 0$.
- (b) There are nonnegative $\mu^{\alpha, p}$ -summable functions H, H_j such that

$$\lim_{j \rightarrow \infty} \int H_j d\mu^{\alpha, p} = 0 \quad \text{and} \quad G(f_j^* - f^*, t, p) \leq H(t) + H_j(t).$$

Indeed, there is the trivial inequality

$$G(f_j^* - f^*, t, p) \leq H_j(t) + [G(f_j^* - f^*, t, p) - H_j(t)]_+,$$

and therefore from (b),

$$\limsup_{j \rightarrow \infty} \int G(f_j^* - f^*, t, p) d\mu^{\alpha, p} t \leq \limsup_{j \rightarrow \infty} \int [\cdots]_+ d\mu^{\alpha, p}.$$

This last expression equals zero since $[\cdots]_+ \rightarrow 0$ pointwise by (a) and also $[\cdots]_+ \leq H \in L^1$.

Part (a). We shall use the inequality $(q+r)^p \leq 2^{p-1}(q^p + r^p)$ for all $q, r \geq 0$. Therefore, $|a-b|^p \leq (|a|+|b|)^p \leq 2^{p-1}(|a|^p + |b|^p)$ for all $a, b \in \mathbb{C}$. With $\delta_j(x) = f_j^*(x) - f^*(x)$, we have $|\delta_j(x) - \delta_j(y)|^p \leq 2^{p-1}(|\delta_j(x)|^p + |\delta_j(y)|^p)$. We also have that $\int \exp[-|x-y|^2 t] d\mathcal{L}^n y$ is independent of x . Inserting this information into the definition of $G(f_j^* - f^*, t, p)$ and using the aforementioned fact that $\int |\delta_j|^p d\mathcal{L}^n \rightarrow 0$, we have the desired conclusion (a) above.

Part (b). Using the inequality again,

$$|\delta_j(x) - \delta_j(y)|^p \leq 2^{p-1}(|f_j^*(x) - f_j^*(y)|^p + |f^*(x) - f^*(y)|^p),$$

so

$$\begin{aligned} G(f_j^* - f^*, t, p) &\leq 2^{p-1}[G(f_j^*, t, p) + G(f^*, t, p)] \\ &\leq 2^{p-1}[G(f_j, t, p) + G(f, t, p)]. \end{aligned}$$

But by the same inequality,

$$G(f_j, t, p) \leq 2^{p-1}[G(f_j - f, t, p) + G(f, t, p)],$$

and thus $G(f_j^* - f^*, t, p) \leq H_j(t) + H(t)$ with $H_j(t) = 2^{2p-2}G(f_j - f, t, p)$ and $H(t) = (2^{2p-2} + 2^{p-1})G(f, t, p)$. By assumption, $\int H_j d\mu^{\alpha, p} \rightarrow 0$, as required in (b) above. ■

REFERENCES

- [AF] F. J. Almgren, *Spherical symmetrization*, Proc. Internat. Workshop on Integral Functionals in the Calculus of Variations, Trieste, September 1985, *Supplemento Rend. Circ. Mat. Palermo* (2) **15** (1987), 1–25.
- [AL1] F. J. Almgren and E. H. Lieb, *The (non)continuity of symmetric decreasing rearrangement*. This is a summary of our results. It is to be published in *Sympos. Math.*, Proc. Conf. on Geometry of Solutions to PDE, Academic Press, New York, and in *Proc. Workshop on Variational Problems*, Paris, June 1988.
- [AL2] ———, *Symmetric decreasing rearrangement can be discontinuous*, *Bull. Amer. Math. Soc.* **20** (1989), 177–180.
- [AAT] A. Ambrosetti and R. E. L. Turner, *Some discontinuous variational problems*, *Differential Integral Equations* **1** (1988), 341–349.
- [AT] T. Aubin, *Problèmes isopérimétriques et espaces de Sobolev*, *C. R. Acad. Sci. Paris Sér. I Math.* **280** (1975), 279–281. See also *J. Differential Geom.* **11** (1976), 573–598.
- [BC] C. Bandle, *Isoperimetric inequalities and applications*, Pitman, Boston, 1980.
- [BBT] J. L. Bona, D. K. Bose, and R. E. L. Turner, *Finite amplitude steady waves in stratified fluids*, *J. Math. Pures Appl.* (9) **62** (1983), 389–439.
- [BL] H. Brézis and E. H. Lieb, *A relation between pointwise convergence of functions and convergence of functionals*, *Proc. Amer. Math. Soc.* **88** (1983), 486–490.
- [BLL] H. J. Brascamp, E. H. Lieb, and J. M. Luttinger, *A general rearrangement inequality for multiple integrals*, *J. Funct. Anal.* **17** (1974), 227–237.
- [BZ] J. Brothers and W. P. Ziemer, *Minimal rearrangements of Sobolev functions*, *J. Reine Angew. Math.* **384** (1988), 153–179.
- [CG] G. Chiti, *Rearrangements of functions and convergence in Orlicz spaces*, *Appl. Anal.* **9** (1979), 23–27.
- [CJ] J.-M. Coron, *The continuity of rearrangement in $W^{1,p}(\mathbf{R})$* , *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **11** (1984), 57–85.
- [CRZ] J. A. Crowe, P. C. Rosenbloom, and J. A. Zweibel, *Rearrangements of functions*, *J. Funct. Anal.* **66** (1986), 432–438.
- [CT] M. G. Crandall and L. Tartar, *Some relations between nonexpansive and order preserving mappings*, *Proc. Amer. Math. Soc.* **78** (1980), 358–390.
- [DG] G. F. D. Duff, *A general inequality for the derivative of an equimeasurable rearrangement*, *Canad. J. Math.* **28** (1976), 793–804.
- [FH] H. Federer, *Geometric measure theory*, Springer-Verlag, New York, 1969.
- [GM] M. Giaquinta, *Multiple integrals in the calculus of variations and nonlinear elliptic systems*, *Ann. of Math. Stud.*, no. 105, Princeton Univ. Press, Princeton, NJ, 1983.
- [HK] K. Hilden, *Symmetrization of functions in Sobolev spaces and the isoperimetric inequality*, *Manuscripta Math.* **18** (1976), 215–235.
- [HLP] G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge Univ. Press, Cambridge, 1934.
- [KB] B. Kawohl, *Rearrangements and convexity of level sets in PDE*, *Lecture Notes in Math.*, no. 1150, Springer-Verlag, New York, 1985.
- [LE1] E. H. Lieb, *Existence and uniqueness of the minimizing solution of Choquard's non-linear equation*, *Stud. Appl. Math.* **57** (1977), 93–105.
- [LE2] ———, *Sharp constants in the Hardy-Littlewood-Sobolev and related inequalities*, *Ann. Math.* **118** (1983), 349–374.
- [PS] G. Polya and G. Szegő, *Isoperimetric inequalities in mathematical physics*, *Ann. of Math. Stud.*, no. 27, Princeton Univ. Press, Princeton, NJ, 1952.
- [RF] F. Riesz, *Sur une inégalité intégrale*, *J. London Math. Soc.* **5** (1930), 162–168.

- [RS] B. Ruf and S. Solimini, *On a class of superlinear Sturm-Liouville problems with arbitrarily many solutions*, Siam J. Math. Anal. **17** (1986), 761–771.
- [SE1] E. Sperner, Jr., *Symmetrisierung für Funktionen mehrerer reeller Variablen*, Manuscripta Math. **11** (1974), 159–170.
- [SE2] —, *Zur Symmetrisierung von Funktionen auf Sphären*, Math. Z. **134** (1973), 317–327.
- [SJ1] J. Serrin, *On a fundamental theorem in the calculus of variations*, Acta Math. **102** (1959), 1–32.
- [SJ2] —, *On the definition and properties of certain variational integrals*, Trans. Amer. Math. Soc. **101** (1961), 139–167.
- [TA] J. E. Taylor and F. J. Almgren, *Optimal crystal shapes*, Variational Methods for Free Surface Interfaces (P. Concus and R. Finn, eds.), Springer-Verlag, New York, 1987, pp. 1–11.
- [TG] G. Talenti, *Best constant in Sobolev inequality*, Ann. Mat. Pura Appl. **110** (1976), 353–372.
- [WH] H. Whitney, *A function not constant on a connected set of critical points*, Duke Math. J. **1** (1935), 514–517.

ABSTRACT. This paper deals with the operation $\mathcal{R}$ of symmetric decreasing rearrangement which maps $\mathbf{W}^{1,p}(\mathbf{R}^n)$ to $\mathbf{W}^{1,p}(\mathbf{R}^n)$. We show that even though it is norm decreasing, $\mathcal{R}$ is not continuous for $n \geq 2$. The functions at which $\mathcal{R}$ is continuous are precisely characterized by a new property called *co-area regularity*. Every sufficiently differentiable function is co-area regular, and both the regular and the irregular functions are dense in $\mathbf{W}^{1,p}(\mathbf{R}^n)$. Curiously, $\mathcal{R}$ is always continuous in fractional Sobolev spaces $\mathbf{W}^{\alpha,p}(\mathbf{R}^n)$ with $0 < \alpha < 1$.

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY 08544