

# RECTIFIABLE-REIFENBERG AND THE REGULARITY OF STATIONARY AND MINIMIZING HARMONIC MAPS

AARON NABER AND DANIELE VALTORTA

ABSTRACT. In this paper we study the regularity of stationary and minimizing harmonic maps  $f : B_2(p) \subseteq M \rightarrow N$  between Riemannian manifolds. If  $S^k(f) \equiv \{x \in M : \text{no tangent map at } x \text{ is } k+1\text{-symmetric}\}$  is the  $k^{\text{th}}$ -stratum of the singular set of  $f$ , then it is well known that  $\dim S^k \leq k$ , however little else about the structure of  $S^k(f)$  is understood in any generality. Our first result is for a general stationary harmonic map, where we prove that  $S^k(f)$  is  $k$ -rectifiable. In fact, we prove for  $k$ -a.e. point  $x \in S^k(f)$  that there exists a unique  $k$ -plane  $V^k \subseteq T_x M$  such that every tangent map at  $x$  is  $k$ -symmetric with respect to  $V$ .

In the case of minimizing harmonic maps we go further, and prove that the singular set  $S(f)$ , which is well known to satisfy  $\dim S(f) \leq n - 3$ , is in fact  $n - 3$ -rectifiable with uniformly finite  $n - 3$ -measure. An effective version of this allows us to prove that  $|\nabla f|$  has estimates in  $L^3_{\text{weak}}$ , an estimate which is sharp as  $|\nabla f|$  may not live in  $L^3$ . More generally, we show that the regularity scale  $r_f$  also has  $L^3_{\text{weak}}$  estimates.

The above results are in fact just applications of a new class of estimates we prove on the quantitative stratifications  $S^k_{\epsilon,r}(f)$  and  $S^k_{\epsilon}(f) \equiv S^k_{\epsilon,0}(f)$ . Roughly,  $S^k_{\epsilon} \subseteq M$  is the collection of points  $x \in M$  for which no ball  $B_r(x)$  is  $\epsilon$ -close to being  $k+1$ -symmetric. We show that  $S^k_{\epsilon}$  is  $k$ -rectifiable and satisfies the Minkowski estimate  $\text{Vol}(B_r S^k_{\epsilon}) \leq Cr^{n-k}$ .

The proofs require a new  $L^2$ -subspace approximation theorem for stationary harmonic maps, as well as new  $W^{1,p}$ -Reifenberg and rectifiable-Reifenberg type theorems. These results are generalizations of the classical Reifenberg, and give checkable criteria to determine when a set is  $k$ -rectifiable with uniform measure estimates. The new Reifenberg type theorems may be of some independent interest. The  $L^2$ -subspace approximation theorem we prove is then used to help break down the quantitative stratifications into pieces which satisfy these criteria.

## CONTENTS

1. Introduction	3
1.1. Results for Stationary Harmonic Maps	6
1.2. Results for Minimizing Harmonic Maps	8
1.3. Results under Additional Hypothesis	9
1.4. Outline of Proofs and Techniques	10
2. Preliminaries	13
2.1. Stationary Harmonic Maps and Monotonicity	13
2.2. Quantitative 0-Symmetry and Cone Splitting	14
2.3. Defect Measure	15
2.4. $\epsilon$ -regularity for Minimizers	16
2.5. Hausdorff, Minkowski, and packing Content	17

---

Date: April 23, 2017.

The first author has been supported by NSF grant DMS-1406259, the second author has been supported by SNSF grant 149539.

2.6. The Classical Reifenberg Theorem	19
2.7. $W^{1,p}$ maps and rectifiability	20
3. The $W^{1,p}$ -Reifenberg and rectifiable-Reifenberg Theorems	20
3.1. Statement of Main $W^{1,p}$ -Reifenberg and rectifiable-Reifenberg Results	21
4. Technical Constructions toward New Reifenberg Results	24
4.1. Hausdorff distance and subspaces	24
4.2. Distance between $L^2$ best planes	26
4.3. Comparison between $L^2$ and $L^\infty$ planes	30
4.4. bi-Lipschitz equivalences	30
4.5. Pointwise Estimates on $D$	35
5. Proof of Theorem 3.3: The Rectifiable-Reifenberg	35
5.1. Weak Upper bound $\lambda^k(S \cap B_r(x)) \leq C(n)r^k$	35
5.2. First induction: upwards	37
5.3. Second induction: downwards: outline	37
5.4. Second induction: downwards: detailed proof	39
5.5. Map and manifold structure.	42
5.6. Volume estimates on the manifold part	44
5.7. Estimates on the excess set	46
5.8. Completion of the weak upper bounds	46
5.9. Rectifiability, $W^{1,p}$ and improved measure bounds	47
6. Proof of Theorems 3.2 and 3.4	53
6.1. Proof of Theorem 3.2: The $W^{1,p}$ -Reifenberg	53
6.2. Proof of Theorem 3.4: The Discrete Rectifiable-Reifenberg	57
7. $L^2$ -Best Approximation Theorems	57
7.1. Symmetry and Gradient Bounds	58
7.2. Best $L^2$ -Subspace Equations	59
7.3. Proof of Theorem 7.1	62
8. The Inductive Covering Lemma	63
8.1. Technical Constructions for Decomposing $U'_r \cup U'_+$	65
8.2. Estimating $U_>$ in Lemma 8.1	69
8.3. Estimating $U_0$ in Lemma 8.1	71
9. Proof of Main Theorems for Stationary Harmonic Maps	73
9.1. Proof of Theorem 1.3	73
9.2. Proof of Theorem 1.4	75
9.3. Proof of Theorem 1.5	78
10. Proof of Main Theorems for Minimizing Harmonic Maps	78
10.1. Proof of Theorem 1.6	78
10.2. Proof of Theorem 1.8	79
11. Sharpness of the results	79
11.1. Sharpness of $L^p$ Estimates for Minimizers	79
11.2. Rectifiable-Reifenberg Example I	79

11.3. Rectifiable-Reifenberg Example II	80
Acknowledgments	80
References	80

## 1. INTRODUCTION

Let  $f : B_2(p) \subseteq M \rightarrow N$  be a stationary harmonic map between two Riemannian manifolds with  $N$  compact without boundary. We recall that the singular set of  $f$  is defined as the complement of the regular set, i.e.,

$$S(f) = \{x \in M \text{ s.t. } \exists r > 0 \text{ s.t. } f|_{B_r(x)} \text{ is continuous}\}^C = \{x \in M \text{ s.t. } \exists r > 0 \text{ s.t. } f|_{B_r(x)} \text{ is smooth}\}^C. \quad (1.1)$$

One can use  $f$  to give a stratification of  $M$  by the singular sets

$$S^0(f) \subseteq \cdots \subseteq S^k(f) \subseteq \cdots \subseteq S(f) \subseteq M, \quad (1.2)$$

defined by

$$S^k(f) \equiv \{x \in M : \text{no tangent map at } x \text{ is } k+1\text{-symmetric}\}, \quad (1.3)$$

see Definition 1.2 for a precise definition and more detailed discussion. A key result of [SU82] is that we have the Hausdorff dimension estimate

$$\dim S^k \leq k, \quad (1.4)$$

However, little else is known about the structure of the singular sets  $S^k(f)$ . In the stationary case, nothing is known in general. In the case of a minimizing harmonic map, [SU82] proves that  $S(f) = S^{n-3}(f)$ , and, under the additional assumption that  $N$  is analytic, it follows from the wonderful work of [Sim96] that the top stratum  $S^{n-3}$  is rectifiable. In low dimensions and when the targets are spheres, these results have been further refined in [LW06] and [HL90].

The goal of this paper is to study additional properties of the stratification and the associated quantitative stratification for stationary and minimizing harmonic maps. That is, a first goal is to see for a stationary harmonic map with a general compact target space that  $S^k(f)$  is  $k$ -rectifiable *for all*  $k$ . In fact, the proof allows us to see the following stronger result. For  $k$ -a.e.  $x \in S^k$  there will exist a unique  $k$ -plane  $V^k \subseteq T_x M$  such that *every* tangent map at  $x$  will be  $k$ -symmetric with respect to  $V$ . Let us observe a subtlety to this statement. We are not claiming that the tangent maps at such  $x$  are unique. We are simply claiming that for  $k$ -a.e. point there is a maximal dimensional  $k$ -plane of symmetry which exists, and that this plane of symmetry is unique independent of which tangent map is chosen. Theorem 1.5 gives the precise results for the classical stratification of a stationary harmonic map.

For minimizing harmonic maps we can improve on this result in two ways. First, we can show that the top stratum is not only  $n - 3$  rectifiable, but has an *a priori* bound on its  $n - 3$  measure. That is, if

$$\int_{B_2(p)} |\nabla f|^2 \leq \Lambda, \quad (1.5)$$

is the  $L^2$  bound on the energy, then we have the  $n - 3$ -Hausdorff measure estimate

$$\lambda^{n-3}(S^{n-3}(f) \cap B_1) \leq C, \quad (1.6)$$

where  $C$  depends on  $M$ ,  $N$  and  $\Lambda$  (see Sections 1.1 and following for precise statements).

Indeed, we show a much stronger Minkowski version of this estimate, see Theorem 1.6 for the complete statement. In fact, we can prove what turns out to be a much more effective analytic version of the above. Namely, in Theorem 1.8 we show that  $|\nabla f|$ , and in fact the regularity scale  $r_f$ , have *a priori* weak  $L^3$  bounds. That is,

$$\text{Vol}(\{|\nabla f| > r^{-1}\} \cap B_1) \leq \text{Vol}(B_r \{|\nabla f| > r^{-1}\} \cap B_1) \leq Cr^3, \quad (1.7)$$

where  $B_r(E)$  stands for the tubular neighborhood of radius  $r$  around the set  $E$ , i.e.,

$$B_r(E) = \{x \text{ s.t. } d(x, E) < r\} = \bigcup_{x \in E} B_r(x). \quad (1.8)$$

Notice that this is indeed sharp, in that there are counterexamples showing that  $|\nabla f| \notin L^3$ , see Example 11.1. Let us also point out that this sharpens results from [CN13b], where it was proven for minimizing harmonic maps that  $|\nabla f| \in L^p$  for all  $p < 3$ . We refer the reader to Section 1.2 for the precise and most general statements.

Now the techniques of this paper all center around estimating the *quantitative* stratification, not the standard stratification itself. In fact, it is for the quantitative stratification that the most important results of the paper hold, everything else can be seen to be corollaries of these statements.

A first version of the quantitative stratification can be found in [Alm00, section 2.25]. This concept was later developed in [CN13a] with the goal of proving new estimates on noncollapsed manifolds with Ricci curvature bounded below and in particular Einstein manifolds, and then extended in [CN13b] to give effective and  $L^p$  estimates on minimal and stationary harmonic maps and minimal submanifolds. It has since been used in [CHN13b], [CHN13a], [CNV15], [FMS15], [BL15] to prove similar results in the areas of mean curvature flow, critical sets of elliptic equations, harmonic map flow, and biharmonic maps.

Before describing the results in this paper on the quantitative stratification, let us give more precise definitions of everything. To begin with, to describe the stratification and quantitative stratification we need to discuss the notion of symmetry associated to a solution. Specifically:

**Definition 1.1.** We define the following:

- (1) A mapping  $f : \mathbb{R}^n \rightarrow N$  is called  $k$ -symmetric if  $f(\lambda x) = f(x) \forall \lambda > 0$  and if there exists a  $k$ -plane  $V^k \subseteq \mathbb{R}^n$  such that for each  $x \in \mathbb{R}^n$  and  $y \in V^k$  we have that  $f(x+y) = f(x)$ . Note that a 0-symmetric function is simply a homogeneous function of degree 0.

- (2) Given a mapping  $f : M \rightarrow N$  and  $\epsilon > 0$ , we say a ball  $B_r(x) \subseteq M$  is  $(k, \epsilon)$ -symmetric if  $r < \text{inj}(x)$  and there exists a  $k$ -symmetric mapping  $\tilde{f} : T_x M \rightarrow N$  such that  $\int_{B_r(x)} |f - \tilde{f}|^2 dv_g < \epsilon$ , where we have used the exponential map to identify  $\tilde{f}$  as a function on  $M$ .

Thus, a function is  $k$ -symmetric if it only depends on  $n - k$  variables and is radially invariant. A  $k$ -symmetric function can therefore be naturally identified with a function on a  $n - k - 1$  sphere. A function is  $(k, \epsilon)$ -symmetric on a ball  $B_r(x)$  if it is  $L^2$  close to a  $k$ -symmetric function on this ball.

With the notion of symmetry in hand we can define precisely the quantitative stratification associated to a solution. The idea is to group points together based on the amount of symmetry that balls centered at those points contain. In fact, there are several variants which will play a role for us, all of which are important for the applications to the standard singular set. Let us introduce them and briefly discuss them:

**Definition 1.2.** For a stationary harmonic map  $f : B_2(p) \rightarrow N$  we make the following definitions:

- (1) For  $\epsilon, r > 0$  we define the  $k^{\text{th}}$   $(\epsilon, r)$ -stratification  $S_{\epsilon, r}^k(f)$  by

$$S_{\epsilon, r}^k(f) \equiv \{x \in B_1(p) : \text{for no } r \leq s < 1 \text{ is } B_s(x) \text{ a } (k+1, \epsilon)\text{-symmetric ball}\}. \quad (1.9)$$

- (2) For  $\epsilon > 0$  we define the  $k^{\text{th}}$   $\epsilon$ -stratification  $S_\epsilon^k(f)$  by

$$S_\epsilon^k(f) = \bigcap_{r>0} S_{\epsilon, r}^k(f) \equiv \{x \in B_1(p) : \text{for no } 0 < r < 1 \text{ is } B_r(x) \text{ a } (k+1, \epsilon)\text{-symmetric ball}\}. \quad (1.10)$$

- (3) We define the  $k^{\text{th}}$ -stratification  $S^k(f)$  by

$$S^k(f) = \bigcup_{\epsilon>0} S_\epsilon^k = \{x \in B_1(p) : \text{no tangent cone at } x \text{ is } k+1\text{-symmetric}\}. \quad (1.11)$$

*Remark 1.1.* It is a small but important exercise to check that the standard stratification  $S^k(f)$  as defined in (1.2) agrees with the set  $\bigcup_{\epsilon>0} S_\epsilon^k$ . We do this carefully in Section 9.3.

Let us discuss in words the meaning of the quantitative stratification, and how it relates to the standard stratification. As discussed at the beginning of the section, the stratification  $S^k(f)$  of  $M$  is built by separating points of  $M$  based on the infinitesimal symmetries of  $f$  at those points. The quantitative stratifications  $S_\epsilon^k(f)$  and  $S_{\epsilon, r}^k(f)$  are, on the other hand, instead built by separating points of  $M$  based on how many symmetries exist on balls of definite size around the points. In practice, the quantitative stratification has two advantages to the standard stratification. First, for applications to minimizing harmonic maps the quantitative stratification allows one to prove effective estimates. In particular, in [CN13b] the weaker  $L^p$  estimates

$$\int_{B_1(p)} |\nabla f|^{3-\delta}, \int_{B_1(p)} |\nabla^2 f|^{\frac{3-\delta}{2}} < C_\delta \forall \delta > 0, \quad (1.12)$$

on solutions were obtained by exploiting this fact. The second advantage is that the estimates on the quantitative stratification are much stronger than those on the standard stratification. Namely, in [CN13b] the Hausdorff dimension estimate (1.4) on  $S^k(f)$  was improved to the Minkowski content estimate

$$\text{Vol}(B_r S_{\epsilon, r}^k) \leq C_\delta r^{n-k-\delta} \forall \delta > 0. \quad (1.13)$$

One of the key technical estimates of this paper is that in Theorem 1.3 we drop the  $\delta$  from the above estimate and obtain an estimate of the form

$$\text{Vol}(B_r S_{\epsilon,r}^k) \leq Cr^{n-k}. \quad (1.14)$$

From this we are able conclude in Theorem 1.4 an estimate on  $S_\epsilon^k$  of the form

$$\text{Vol}(B_r S_\epsilon^k) \leq Cr^{n-k}. \quad (1.15)$$

In particular, this estimate allows us to conclude that  $S_\epsilon^k$  has uniformly finite  $k$ -dimensional measure. In fact, the techniques will prove much more for us. They will show us that  $S_\epsilon^k$  is  $k$ -rectifiable, and that for  $k$ -a.e. point  $x \in S_\epsilon^k$  there is a unique  $k$ -plane  $V^k \subseteq T_x M$  such that *every* tangent map at  $x$  is  $k$ -symmetric with respect to  $V$ . By observing that  $S^k(f) = \bigcup S_\epsilon^k(f)$ , this allows us to prove in Theorem 1.5 our main results on the classical stratification. This decomposition of  $S^k$  into the pieces  $S_\epsilon^k$  is crucial for the proof.

On the other hand, (1.14), combined with the  $\epsilon$ -regularity theorems of [SU82],[CN13b], allow us to conclude in the minimizing case both the weak  $L^3$  estimate on  $|\nabla f|$ , and the  $n - 3$ -finiteness of the singular set of  $f$ . Thus we will see that Theorems 1.6 and 1.8 are fairly quick consequences of (1.15).

Thus we have seen that (1.14) and (1.15), and more generally Theorem 1.3 and Theorem 1.4, are the main challenges of the paper. We will give a more complete outline of the proof in Section 1.4, however let us mention for the moment that two of the new ingredients to the proof are a new  $L^2$ -subspace approximation theorem for stationary harmonic maps, proved in Section 7, and new  $W^{1,p}$ -Reifenberg and rectifiable-Reifenberg theorems, proved in Section 3. The  $L^2$ -approximation result roughly tells us that the  $L^2$ -distance of a measure from being contained in a  $k$ -dimensional subspace may be estimated by integrating the energy drop of a stationary harmonic map over the measure. To exploit the estimate we prove a new  $W^{1,p}$ -Reifenberg type theorem. The classical Reifenberg theorem states that if we have a set  $S$  which is  $L^\infty$ -approximated by a subspace at every point and scale, then  $S$  is bi-Hölder to a manifold, see theorem 2.18 for a precise statement. It is important for us to improve on this bi-Hölder estimate, at least enough that we are able to control the  $k^{\text{th}}$ -dimensional measure of the set and prove rectifiability. In particular, we want to improve the  $C^\alpha$ -maps to  $W^{1,p}$ -maps for  $p > k$ , and we will want to do it using a condition which is integral in nature. More precisely, we will only require a form of summable  $L^2$ -closeness of the subset  $S$  to the approximating subspaces. We will see in Theorem 7.1 that by using the  $L^2$ -subspace approximation theorem that the conditions of this new rectifiable-Reifenberg are in fact controllable for the quantitative stratifications  $S_\epsilon^k$ , as least after we break it up into appropriate pieces.

**1.1. Results for Stationary Harmonic Maps.** We now turn our attention to giving precise statements of the main results of this paper. In this subsection we focus on those concerning the singular structure of stationary harmonic maps. That is, we will be considering stationary harmonic maps  $f : B_2(p) \subseteq M \rightarrow N$ . In order to write the constants involved in some explicit form, let us choose  $K_M, K_N, n > 0$  to be the smallest

number such that

$$\begin{aligned} |\sec_{B_2(p)}| &\leq K_M, \quad \text{inj}(B_2(p)) \geq K_M^{-1}, \\ |\sec_N| &\leq K_N, \quad \text{inj}(N) \geq K_N^{-1}, \quad \text{diam}(N) \leq K_N, \\ \dim(M), \dim(N) &\leq n. \end{aligned} \tag{1.16}$$

Now let us begin by discussing our main theorem for the quantitative stratifications  $S_{\epsilon,r}^k(f)$ :

**Theorem 1.3** ( $(\epsilon, r)$ -Stratification of Stationary Harmonic Maps). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a stationary harmonic mapping satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Then for each  $\epsilon > 0$  there exists  $C_\epsilon(n, K_M, K_N, \Lambda, \epsilon)$  such that*

$$\text{Vol}(B_r(S_{\epsilon,r}^k(f))) \leq C_\epsilon r^{n-k}. \tag{1.17}$$

When we study the stratum  $S_\epsilon^k(f)$ , we can refine the above to prove structure theorems on the set itself. For the definition of  $k$ -rectifiability, we refer the reader to the standard reference [Fed69].

**Theorem 1.4** ( $\epsilon$ -Stratification of Stationary Harmonic Maps). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a stationary harmonic mapping satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Then for each  $\epsilon > 0$  there exists  $C_\epsilon(n, K_M, K_N, \Lambda, \epsilon)$  such that*

$$\text{Vol}(B_r(S_\epsilon^k(f))) \leq C_\epsilon r^{n-k}. \tag{1.18}$$

*In particular, we have the  $k$ -dimensional Hausdorff measure estimate  $\lambda^k(S_\epsilon^k(f)) \leq C_\epsilon$ . Further,  $S_\epsilon^k(f)$  is  $k$ -rectifiable, and for  $k$ -a.e.  $x \in S_\epsilon^k$  there exists a unique  $k$ -plane  $V^k \subseteq T_x M$  such that every tangent cone of  $x$  is  $k$ -symmetric with respect to  $V^k$ .*

*Remark 1.2.* In fact, the techniques will prove an estimate even stronger than the Minkowski estimate of (1.17) and (1.18). That is, one can prove a uniform  $k$ -dimensional packing content estimate. More precisely, let  $\{B_{r_i}(x_i)\}$  be any collection of disjoint balls such that  $x_i \in S_\epsilon^k$ , then we have the content estimate  $\sum r_i^k \leq C_\epsilon$ .

Finally, we end this subsection by stating our main results when it comes to the classical stratification  $S^k(f)$ . The following may be proved from the previous Theorem in only a few lines given the formula  $S^k(f) = \bigcup S_\epsilon^k(f)$ :

**Theorem 1.5** (Stratification of Stationary Harmonic Maps). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a stationary harmonic mapping satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Then for each  $k$  we have that  $S^k(f)$  is  $k$ -rectifiable. Further, for  $k$ -a.e.  $x \in S^k(f)$  there exists a unique  $k$ -plane  $V^k \subseteq T_x M$  such that every tangent cone of  $x$  is  $k$ -symmetric with respect to  $V^k$ .*

**1.2. Results for Minimizing Harmonic Maps.** In this section we record our main results for minimizing harmonic maps. Most of the results of this section follow quickly by combining the quantitative stratification results of Section 1.1 with the  $\epsilon$ -regularity of [SU82, CN13b], see Section 2.4 for a review of these points.

Our first estimate is on the singular set  $S(f)$  of a minimizing harmonic map. Recall that  $S(f)$  is the set of points where  $f$  is not smooth. Our first estimate on the singular structure of a minimizing harmonic map is the following:

**Theorem 1.6** (Structure of Singular Set). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a minimizing harmonic mapping satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Then  $S(f)$  is  $n - 3$ -rectifiable and there exists  $C(n, K_M, K_N, \Lambda)$  such that*

$$\text{Vol}(B_r(S(f)) \cap B_1(p)) \leq Cr^3. \quad (1.19)$$

In particular,  $\lambda^{n-3}(S(f)) \leq C$ .

*Remark 1.3.* As in Theorem 1.4, the same techniques prove a packing content estimate on  $S(f)$ . That is, if  $\{B_{r_i}(x_i)\}$  is any collection of disjoint balls such that  $x_i \in S(f) \cap B_1$ , then we have the content estimate  $\sum r_i^k \leq C$ .

The above can be extended to effective Schauder estimates on  $f$ . To state the results in full strength let us recall the notion of the regularity scale associated to a function. Namely:

**Definition 1.7.** Let  $f : B_2(p) \rightarrow N$  be a mapping between Riemannian manifolds. For  $x \in B_1(p)$  we define the regularity scale  $r_f(x)$  by

$$r_f(x) \equiv \max\{0 \leq r \leq 1 : \sup_{B_r(x)} |\nabla f| \leq r^{-1}\}. \quad (1.20)$$

By definition,  $r_f(x) \equiv 0$  if  $f$  is not Lipschitz in a neighborhood of  $x$ .

*Remark 1.4.* The regularity scale has nice scaling property. Indeed, if we define  $T_{X,\rho} : T_x M \rightarrow N$  by

$$T_{x,\rho}(y) = f(\exp_x(\rho y)), \quad (1.21)$$

then  $r_f(x) = \rho r_{T_{x,\rho}}(0)$ . In other words, if  $r \equiv r_f(x)$  and we rescale  $B_r(x) \rightarrow B_1(0)$ , then on the rescaled ball we will have that  $|\nabla T_{x,r}| \leq 1$  on  $B_1(0)$ .

*Remark 1.5.* We have the easy estimate  $|\nabla f|(x) \leq r_f(x)^{-1}$ . However, a lower bound on  $r_f(x)$  is in principle *much* stronger than an upper bound on  $|\nabla f|(x)$ .

*Remark 1.6.* Notice that the regularity scale is a Lipschitz function with  $|\nabla r_f(x)| \leq 1$ .

*Remark 1.7.* If  $f$  satisfies an elliptic equation, e.g. is weakly harmonic, then we have the estimate

$$\sup_{B_{r_f/2}(x)} |\nabla^k f| \leq C_k r_f(x)^{-k}. \quad (1.22)$$

In particular, control on  $r_f$  gives control on all higher order derivatives.



Now let us state our main estimates for minimizing harmonic maps:

**Theorem 1.8** (Estimates on Minimizing Harmonic Maps). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a minimizing harmonic mapping satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Then there exists  $C(n, K_M, K_N, \Lambda)$  such that*

$$\text{Vol}(\{x \in B_1(p) : |\nabla f| > r^{-1}\}) \leq \text{Vol}(\{x \in B_1(p) : r_f(x) < r\}) \leq Cr^3. \quad (1.23)$$

*In particular, both  $|\nabla f|$  and  $r_f^{-1}$  have uniform bounds in  $L^3_{\text{weak}}(B_1(p))$ , the space of weakly  $L^3$  functions. In fact, we also have the Hessian estimate*

$$\text{Vol}(\{x \in B_1(p) : |\nabla^2 f| > r^{-2}\}) \leq Cr^3, \quad (1.24)$$

*which in particular gives us uniform bounds for  $|\nabla^2 f|$  in  $L^{3/2}_{\text{weak}}(B_1(p))$ , the space of weakly  $L^{3/2}$  functions.*

*Remark 1.8.* The  $L^3_{\text{weak}}$  estimates are sharp in that there exist examples for which  $|\nabla f|$  does not live in  $L^3$ , see Section 11.1.

**1.3. Results under Additional Hypothesis.** The previous two subsections have focused on results for completely general stationary or minimizing harmonic maps. In this subsection we would like to see how these results may be improved under further assumptions. Specifically, for stationary harmonic maps  $f : M \rightarrow N$  we would like to see that the regularity results may be improved to match those of minimizing harmonic maps if we assume there are no smooth harmonic maps from  $S^2$  into  $N$ . The idea behind this follows that of [Lin99],[CHN13a]. Additionally, though the regularity results of Sections 1.1 and 1.2 are sharp in complete generality, they may be improved if we assume that  $N$  has no other stationary or minimizing harmonic maps from  $S^k$  into  $N$ . Precisely, the main result of this subsection for stationary harmonic maps is the following:

**Theorem 1.9** (Improved Estimates for Stationary Maps). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a stationary harmonic mapping satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Assume further that for some  $k \geq 2$  there exists no smooth nonconstant stationary harmonic maps  $S^\ell \rightarrow N$  for all  $\ell \leq k$ . Then there exists  $C(n, K_M, N, \Lambda)$  such that*

$$\text{Vol}(\{x \in B_1(p) : |\nabla f| > r^{-1}\}) \leq \text{Vol}(\{x \in B_1(p) : r_f(x) < r\}) \leq Cr^{2+k}. \quad (1.25)$$

*In particular, both  $|\nabla f|$  and  $r_f^{-1}$  have uniform bounds in  $L^{2+k}_{\text{weak}}(B_1(p))$ , the space of weakly  $L^{2+k}$  functions.*

*Remark 1.9.* The proof of Theorem 1.9 follows verbatim the proof of Theorem 1.8, except one replaces the  $\epsilon$ -regularity of theorem 2.8 with the  $\epsilon$ -regularity of theorem 2.11.

In the context where  $f$  is minimizing we have a similar improvement, though in this case we only need to assume there exists no minimizing harmonic maps from  $S^k \rightarrow N$ :

**Theorem 1.10** (Improved Estimates for Minimizing Maps). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a minimizing harmonic mapping satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Assume that for some  $k \geq 2$  there exists no smooth nonconstant minimizing harmonic maps  $S^\ell \rightarrow N$  for all  $\ell \leq k$ . Then there exists  $C(n, K_M, N, \Lambda)$  such that*

$$\text{Vol}(\{x \in B_1(p) : |\nabla f| > r^{-1}\}) \leq \text{Vol}(\{x \in B_1(p) : r_f(x) < r\}) \leq Cr^{2+k}. \quad (1.26)$$

*In particular, both  $|\nabla f|$  and  $r_f^{-1}$  have uniform bounds in  $L_{\text{weak}}^{2+k}(B_1(p))$ , the space of weakly  $L^{2+k}$  functions.*

*Remark 1.10.* The proof of Theorem 1.10 follows verbatim the proof of Theorem 1.8, except one replaces the  $\epsilon$ -regularity of theorem 2.8 with the  $\epsilon$ -regularity of theorem 2.10.

**1.4. Outline of Proofs and Techniques.** In this subsection we give a brief outline of the proof of the main theorems. To describe the new ingredients involved it will be helpful to give a comparison to the proofs of previous results in the area, in particular the dimension estimate (1.4) of [SU82] (which are similar in spirit to the estimates for minimal surfaces in [Alm68] and [Alm00, theorem 2.26]) and the Minkowski and  $L^p$  estimates (1.12) of [CN13b].

Indeed, the starting point for the study of singular sets for solutions of geometric equations typically looks the same, that is, one needs a monotone quantity. In the case of harmonic maps  $f : M \rightarrow N$  between Riemannian manifolds we consider the normalized Dirichlet energy

$$\theta_r(x) \equiv r^{2-n} \int_{B_r(x)} |\nabla f|^2. \quad (1.27)$$

For simplicity sake let us take  $M \equiv \mathbb{R}^n$  in this discussion, which is really of no loss except for some small technical work (it is the nonlinearity of  $N$  that is the difficulty). Then  $\frac{d}{dr}\theta_r \geq 0$ , and  $\theta_r(x)$  is independent of  $r$  if and only if  $f$  is 0-symmetric, see Section 2.1 for more on this. Interestingly, this is the only information one requires to prove the dimension estimate (1.4). Namely, since  $\theta_r(x)$  is monotone and bounded, it must converge as  $r$  tends to zero. In particular, if we consider the sequence of scales  $r_\alpha = 2^{-\alpha}$  then we have for each  $x$  that

$$\lim_{\alpha \rightarrow \infty} \left| \theta_{r_\alpha}(x) - \theta_{r_{\alpha+1}}(x) \right| \rightarrow 0. \quad (1.28)$$

From this one can conclude that *every* tangent map of  $f$  is 0-symmetric. This fact combined with some very general dimension reduction arguments originating with Federer from geometric measure theory [Sim83], which do not depend on the harmonic behavior of  $f$  at all, yield the dimension estimate (1.4) from [SU82].

The improvement in [CN13b] of the Hausdorff dimension estimate (1.4) to the Minkowski content estimate (1.13), and therefore the  $L^p$  estimate of (1.12), requires exploiting more about the monotone quantity  $\theta_r(x)$  than that it limits as  $r$  tends to zero. Indeed, an effective version of (1.28) says that for each  $\delta > 0$  there exists  $N(\Lambda, \delta) > 0$  such that

$$\left| \theta_{r_\alpha}(x) - \theta_{r_{\alpha+1}}(x) \right| < \delta, \quad (1.29)$$

holds for all except for at most  $N$  scales  $\alpha \in \{\alpha_1, \dots, \alpha_N\} \subseteq \mathbb{N}$ . These *bad* scales where (1.29) fails may differ from point to point, but the number of such scales is uniformly bound. This allows one to conclude

that for all but at most  $N$ -scales that  $B_{r_\alpha}(x)$  is  $(0, \epsilon)$ -symmetric, see Section 2.2. To exploit this information a new technique other than dimension reduction was required in [CN13b]. Indeed, in [CN13b] the quantitative 0-symmetry of (1.29) was instead combined with the notion of cone splitting and an energy decomposition in order to conclude the estimates (1.12),(1.13). Since we will use them in this paper, the quantitative 0-symmetry and cone splitting will be reviewed further in Section 2.2.

Now let us begin to discuss the results of this paper. The most challenging aspect of this paper is the proof of the estimates on the quantitative stratifications of Theorems 1.3 and 1.4, and so we will focus on these in our outline. Let us first observe that it might be advantageous to replace (1.29) with a version that forces an actual rate of convergence, see for instance [Sim96]. More generally, if one is in a context where an effective version of tangent cone uniqueness can be proved then this may be exploited. In fact, in the context of critical sets of elliptic equations one can follow exactly this approach, see the authors work [NV] where versions of Theorems 1.3 and 1.4 were first proved in this context. However, in the general context of this paper such an approach fails, as tangent cone uniqueness is not available, and potentially not correct.

Instead, we will first replace (1.29) with the following relatively simple observation. Namely, for each  $x$  there exists  $N(\Lambda, \delta)$  and a finite number of scales  $\{\alpha_1, \dots, \alpha_N\} \subseteq \mathbb{N}$  such that

$$\sum_{\alpha_j < \alpha < \alpha_{j+1}} \left| \theta_{r_\alpha}(x) - \theta_{r_{\alpha+1}}(x) \right| < \delta. \quad (1.30)$$

That is, not only does the energy drop by less than  $\delta$  between these scales, but the sum of all the energy drops is less than  $\delta$  between these scales.

Unfortunately, exploiting (1.30) turns out to be *substantially* harder to use than exploiting either (1.28) or even (1.29). In essence, this is because it is not a local assumption in terms of scale, and one needs estimates which can see many scales simultaneously, but which do not require any form of tangent cone uniqueness statements. Accomplishing this requires two new ingredients, a new rectifiable-Reifenberg type theorem, and a new  $L^2$ -best subspace approximation theorem for stationary harmonic maps, which will allow us to apply the rectifiable-Reifenberg. Let us discuss these two ingredients separately.

We begin by discussing the new  $W^{1,p}$ -Reifenberg and rectifiable-Reifenberg theorems, which are introduced and proved in Section 3. Recall that the classical Reifenberg theorem, reviewed in Section 2.6, gives criteria under which a set becomes  $C^\alpha$ -Hölder equivalent to a ball  $B_1(0^k)$  in Euclidean space. In the context of this paper, it is important to improve on this result so that we have gradient and volume control of our set. Let us remark that there have been many generalizations of the classical Reifenberg theorem in the literature, see for instance [Tor95] [DT12], however those results have hypotheses which are much too strong for the purposes of this paper. Instead, we will focus on improving the  $C^\alpha$ -equivalence to a  $W^{1,p}$ -equivalence. This is strictly stronger by Sobolev embedding, and if  $p > k$  then this results in volume estimates and a rectifiable structure for the set. More generally, we will require a version of the theorem which allows for more degenerate structural behavior, namely a rectifiable-Reifenberg theorem. In this case, the assumptions will conclude that a set  $S$  is rectifiable with volume estimates. Of course, what is key about this result is that the criteria will be checkable for our quantitative stratifications, thus let us discuss these criteria briefly.

Roughly, if  $S \subseteq B_1(0^n)$  is a subset equipped with the  $k$ -dimensional Hausdorff measure  $\lambda^k$ , then let us define the  $k$ -dimensional distortion of  $S$  by

$$D_S^k(y, s) \equiv s^{-2} \inf_{L^k} s^{-k} \int_{S \cap B_s(y)} d^2(z, L^k) d\lambda^k(z), \quad (1.31)$$

where the inf is taken over all  $k$ -dimensional affine subspaces of  $\mathbb{R}^n$ . That is,  $D^k$  measures how far  $S$  is from being contained in a  $k$ -dimensional subspace.

Our rectifiable-Reifenberg then requires this be small on  $S$  in an integral sense, more precisely that

$$r^{-k} \int_{S \cap B_r(x)} \sum_{r_\alpha \leq r} D^k(y, r_\alpha) d\lambda^k(y) < \delta^2. \quad (1.32)$$

For  $\delta$  sufficiently small, the conclusions of the rectifiable-Reifenberg Theorem 3.3 are that the set  $S$  is rectifiable with effective bounds on the  $k$ -dimensional measure. Let us remark that one cannot possibly conclude better than rectifiable under this assumption, see for instance the Examples of Section 11.

Thus, in order to prove the quantitative stratification estimates of Theorems 1.3 and 1.4, we will need to verify that the integral conditions (1.32) hold for the quantitative stratifications  $S_\epsilon^k(f)$ ,  $S_{\epsilon,r}^k(f)$  on all balls  $B_r(x)$ . In actuality the proof is more complicated. We will need to apply a discrete version of the rectifiable-Reifenberg, which will allow us to build an iterative covering of the quantitative stratifications, and each of these will satisfy (1.32). This will allow us to keep effective track of all the estimates involved. However, let us for the moment just focus on the main estimates which allows us to turn (1.32) into information about our harmonic maps, without worrying about such details.

Namely, in Section 7 we prove a new and very general approximation theorem for stationary harmonic maps. As always in this outline, let us assume  $M \equiv \mathbb{R}^n$ , the general case is no harder and we simply work on an injectivity radius ball. Thus we consider a stationary harmonic map  $f : B_{16}(0^n) \rightarrow N$ , as well as an arbitrary measure  $\mu$  which is supported on  $B_1(0^n)$ . We would like to study how closely the support of  $\mu$  can be approximated by a  $k$ -dimensional affine subspace  $L^k \subseteq \mathbb{R}^n$ , in the appropriate sense, and we would like to estimate this distance by using properties of  $f$ . Indeed, if we assume that  $B_8(0^n)$  is *not*  $(k+1, \epsilon)$ -symmetric with respect to  $f$ , then for an arbitrary  $\mu$  we will prove in Theorem 7.1 that

$$\inf_{L^k \subseteq \mathbb{R}^n} \int d^2(x, L^k) d\mu \leq C \int |\theta_8(x) - \theta_1(x)| d\mu, \quad (1.33)$$

where  $C$  will depend on  $\epsilon$ , energy bound of  $f$ , and the geometry of  $M$  and  $N$ . That is, if  $f$  does not have  $k+1$  degrees of symmetry, then how closely the support of an arbitrary measure  $\mu$  can be approximated by a  $k$ -dimensional subspace can be estimated by looking at the energy drop of  $f$  along  $\mu$ . In applications,  $\mu$  will be the restriction to  $B_1$  of some discrete approximation of the  $k$ -dimensional Hausdorff measure on  $S_\epsilon^k$ , and thus the symmetry assumption on  $f$  will hold for all balls centered on the support of  $\mu$ .

In practice, applying (1.33) to (1.32) is subtle and must be done inductively on scale. Additionally, in order to prove the effective Hausdorff estimates  $\lambda^k(S_\epsilon^k \cap B_r) \leq Cr^k$  we will need to use the Covering Lemma 8.1 to break up the quantitative stratification into appropriate pieces, and we will apply the estimates to these. This decomposition is based on a covering scheme first introduced by the authors in [NV]. Thus for the purposes of our outline, let us assume we have already proved the Hausdorff estimate  $\lambda^k(S_\epsilon^k \cap B_r) \leq Cr^k$ ,

and use this to be able to apply the rectifiable-Reifenberg in order to conclude the rectifiability of the singular set. This may feel like a large assumption, however it turns out the proof of the Hausdorff estimate will itself be proved by a very similar tactic, though will require an inductive argument on scale, and will use the discrete rectifiable-Reifenberg of Theorem 3.4 in place of the rectifiable-Reifenberg of Theorem 3.3.

Thus let us choose a ball  $B_r$  and let  $E \equiv \sup_{B_r} \theta_r(y)$ . Let us consider the subset  $\tilde{S}_\epsilon^k \subseteq S_\epsilon^k \cap B_r$  defined by

$$\tilde{S}_\epsilon^k \equiv \{y \in S_\epsilon^k \cap B_r : \theta_0(y) > E - \eta\}, \quad (1.34)$$

where  $\eta = \eta(n, K_M, K_N, \Lambda, \epsilon)$  will be chosen appropriately later. We will show now that  $\tilde{S}_\epsilon^k$  is rectifiable. Since  $\eta$  is fixed and the ball  $B_r$  is arbitrary, the rectifiability of all of  $S_\epsilon^k$  follows quickly from a covering argument. Thus, let us estimate (1.32) by plugging in (1.33) and the Hausdorff estimate to conclude:

$$\begin{aligned} & r^{-k} \int_{\tilde{S}_\epsilon^k} \sum_{r_\alpha \leq r} D^k(x, r_\alpha) d\lambda^k \\ &= r^{-k} \int_{\tilde{S}_\epsilon^k} \sum_{r_\alpha \leq r} \left( \inf_{L^k} r_\alpha^{-2-k} \int_{\tilde{S}_\epsilon^k \cap B_{r_\alpha}(x)} d^2(y, L^k) d\lambda^k(y) \right) d\lambda^k(x) \\ &\leq Cr^{-k} \int_{\tilde{S}_\epsilon^k} \sum_{r_\alpha \leq r} \left( r_\alpha^{-k} \int_{\tilde{S}_\epsilon^k \cap B_{r_\alpha}(x)} |\theta_{8r_\alpha}(y) - \theta_{r_\alpha}(y)| d\lambda^k(y) \right) d\lambda^k(x) \\ &= Cr^{-k} \sum_{r_\alpha \leq r} r_\alpha^{-k} \int_{\tilde{S}_\epsilon^k} \lambda^k(\tilde{S}_\epsilon^k \cap B_{r_\alpha}(y)) |\theta_{8r_\alpha}(y) - \theta_{r_\alpha}(y)| d\lambda^k(y), \\ &\leq Cr^{-k} \int_{\tilde{S}_\epsilon^k} \sum_{r_\alpha \leq r} |\theta_{8r_\alpha}(y) - \theta_{r_\alpha}(y)| d\lambda^k(y), \\ &\leq Cr^{-k} \int_{\tilde{S}_\epsilon^k} |\theta_{8r}(y) - \theta_0(y)| d\lambda^k(y) \\ &\leq Cr^{-k} \lambda_k(\tilde{S}_\epsilon^k) \cdot \eta \\ &< \delta^2. \end{aligned} \quad (1.35)$$

where in the last line we have chosen  $\eta = \eta(n, K_M, K_N, \Lambda, \epsilon)$  so that the estimate is less than the required  $\delta$  from the rectifiable-Reifenberg. Thus we can apply the rectifiable-Reifenberg of Theorem 3.3 in order to conclude the rectifiability of the set  $\tilde{S}_\epsilon^k$ , which in particular proves that  $S_\epsilon^k$  is itself rectifiable, as claimed.

## 2. PRELIMINARIES

**2.1. Stationary Harmonic Maps and Monotonicity.** The key technical tool available to a stationary harmonic map  $f : B_2(p) \subseteq M \rightarrow N$ , which is not available to a weakly harmonic map, is that of a monotone quantity. Namely, given  $x \in B_1(p)$  and  $r \leq 1$  we can consider the normalized Dirichlet energy defined by

$$\theta_r(x) \equiv r^{2-n} \int_{B_r(x)} |\nabla f|^2. \quad (2.1)$$

If  $M \equiv \mathbb{R}^n$ , then for each  $x$  fixed we have that  $\theta_r(x)$  is exactly a monotone increasing function of  $r$ . More precisely, we have that

$$\frac{d}{dr}\theta_r(x) = 2r^{2-n} \int_{\partial B_r(x)} \left| \frac{\partial f}{\partial r} \right|^2 = 2r^{2-n} \int_{\partial B_r(x)} |\langle \nabla f, n_x \rangle|^2. \quad (2.2)$$

Hence, we see that if  $\theta_r(x)$  is independent of  $r$  then we have that  $f$  is 0-symmetric with respect to  $x$ . More generally, if  $\theta_s(x) = \theta_r(x)$  then  $f$  is radially symmetric on the annulus  $A_{s,r}(x) = B_r(x) \setminus B_s(x)$ . Motivated by this, we see that what we are really interested in is the amount the Dirichlet energy drops from one scale to the next, and thus we define

$$W_{s,r}(x) \equiv \theta_r(x) - \theta_s(x) \geq 0. \quad (2.3)$$

Oftentimes we will want to enumerate our choice of scale, so let us define the scales  $r_\alpha \equiv 2^{-\alpha}$  for  $\alpha \geq 0$ , and the corresponding Dirichlet energy drop:

$$W_\alpha(x) \equiv W_{r_\alpha, r_{\alpha-3}}(x) \equiv \theta_{r_{\alpha-3}}(x) - \theta_{r_\alpha}(x) \geq 0. \quad (2.4)$$

In the general case, i.e., when  $M \neq \mathbb{R}^n$ , essentially the same statements may be made, however  $\theta_r(x)$  is now only *almost* monotone, meaning that  $e^{Cr}\theta_r(x)$  is monotone for some constant which depends only on the geometry of  $M$ . From this one can prove that at every point, every tangent map is 0-symmetric, which is the starting point for the dimension estimate (1.4) of [SU82]. In Section 2.2 we will discuss quantitative versions of this point, first introduced in [CN13b] and used in this paper as well, and also generalizations which involve higher degrees of symmetry. These points were first used in [CN13b] to prove Minkowski estimates on the quantitative stratification of a stationary harmonic map. They will also play a role in our arguments, though in a different manner.

**2.2. Quantitative 0-Symmetry and Cone Splitting.** In this subsection we review some of the quantitative symmetry and splitting results of [CN13b], in particular those which will play a role in this paper.

The first result we will discuss acts as an *effective* formulation of the fact that every tangent map is 0-symmetric. Namely, the quantitative 0-symmetry of [CN13b] states that for each  $\epsilon > 0$  and point  $x \in B_1$ , all but a finite number of the balls  $\{B_{r_\alpha}(x)\}_{\alpha \in \mathbb{N}}$  are  $(0, \epsilon)$ -symmetric, where  $r_\alpha \equiv 2^{-\alpha}$ . Precisely:

**Theorem 2.1** (Quantitative 0-Symmetry [CN13b]). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a stationary harmonic map satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$  and let  $\epsilon > 0$  be fixed. Then the following hold:*

- (1) *There exists  $\delta(n, K_N, \Lambda, \epsilon) > 0$  such that for each  $x \in B_1(p)$  if  $|\theta_r(x) - \theta_{\delta r}(x)| < \delta$  and  $K_M < \delta$ , then  $B_r(x)$  is  $(0, \epsilon)$ -symmetric.*
- (2) *For each  $x \in B_1(p)$  there exists a finite number of scales  $\{\alpha_1, \dots, \alpha_K\} \subseteq \mathbb{N}$  with  $K \leq K(n, K_M, K_N, \Lambda, \epsilon)$  such that for  $r \notin (r_{\alpha_j}/2, 2r_{\alpha_j})$  we have that  $B_r(x)$  is  $(0, \epsilon)$ -symmetric.*

*Remark 2.1.* In [CN13b] it is stated that the constant depends on the manifold  $N$ , not just  $K_N$ , however it is easy to check that only  $K_N$  is important in the proof. See the proof of Lemma 7.2 for a relevant argument.

*Remark 2.2.* The assumption  $K_M < \delta$  is of little consequence, since this just means focusing the estimates on balls of sufficiently small radius after rescaling.

Another technical tool that played an important role in [CN13b] was that of cone splitting. This will be used in this paper when proving the existence of unique tangent planes of symmetry for the singular set, so we will discuss it here. In short, cone splitting is the idea that multiple 0-symmetries add to give rise to a  $k$ -symmetry. To state it precisely let us give a careful definition of the notion of independence of a collection of points:

**Definition 2.2.** We say a collection of points  $\{x_1, \dots, x_\ell\} \in \mathbb{R}^n$  is independent if they are linearly independent. We say the collection is  $\tau$ -independent if  $d(x_{k+1}, \text{span}\{x_1, \dots, x_k\}) > \tau$  for each  $k$ . If  $\{x_1, \dots, x_\ell\} \in M$  then we say the collection is  $\tau$ -independent with respect to  $x$  if  $d(x, x_j) < \text{inj}(x)$  and the collection is  $\tau$ -independent when written in exponential coordinates at  $x$ .

Now we are in a position to state the effective cone splitting of [CN13b]:

**Theorem 2.3** (Cone Splitting [CN13b]). *Let  $f : B_3(p) \subseteq M \rightarrow N$  be a stationary harmonic map satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$  and let  $\epsilon, \tau > 0$  be fixed. Then there exists  $\delta(n, K_M, K_N, \Lambda, \epsilon, \tau) > 0$  such that if  $K_M < \delta$  and  $x_0, \dots, x_k \in B_1(p)$  are such that*

- (1)  $B_2(x_j)$  are  $(0, \delta)$ -symmetric,
- (2)  $\{x_0, \dots, x_k\}$  are  $\tau$ -independent at  $p$ ,

then  $B_1(p)$  is  $(k, \epsilon)$ -symmetric.

We end with the following, which one can view as a quantitative form of dimension reduction. The proof is standard and can be accomplished by a contradiction argument:

**Theorem 2.4** (Quantitative Dimension Reduction). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a stationary harmonic map satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Then for each  $\epsilon > 0$  there exists  $\delta(n, K_M, K_N, \Lambda, \epsilon), r(n, K_M, K_N, \Lambda, \epsilon) > 0$  such that if  $B_2(p)$  is  $(k, \delta)$ -symmetric with respect to some  $k$ -plane  $V^k$ , then for each  $x \in B_1(p) \setminus B_\epsilon(V^k)$  we have that  $B_r(x)$  is  $(k+1, \epsilon)$ -symmetric.*

**2.3. Defect Measure.** Since it will play a role for us in one technical aspect of this paper, let us recall in this section the defect measure of a sequence of stationary harmonic maps. We begin with a definition:

**Definition 2.5.** Let  $f_i : B_2(p) \rightarrow N$  be a sequence of stationary harmonic maps satisfying (1.16) with  $\int_{B_2(p)} |\nabla f_i|^2 \leq \Lambda$ . Then, after possibly passing to a subsequence,  $f_i$  has a weak  $W^{1,2}$  limit  $f$ , and we can consider the measure

$$|\nabla f_i|^2 dv_g \rightharpoonup^* |\nabla f|^2 dv_g + \nu, \quad (2.5)$$

where this last convergence is intended in the weak sense of Radon measures. The measure  $\nu$  is nonnegative by Fatou's lemma, and it is called the defect measure associated to  $f_i \rightharpoonup^* f$ .

*Remark 2.3.* We can also allow  $f_i : B_2(p_i) \rightarrow N$  to be defined on different manifolds. In practice, this will occur under the assumption  $K_{M_i} \rightarrow 0$ , so that  $\nu$  becomes a measure on  $B_2(0^n)$ .



The following theorem is one of the main accomplishments of [Lin99]:

**Theorem 2.6** (Rectifiability of Defect Measure). [Lin99, lemma 1.7] *If  $\nu$  is a defect measure as in definition 2.5, then it is  $n - 2$  rectifiable.*

A key tool in the proof of the above, which will be useful in this paper as well, is the following 0-symmetry result:

**Theorem 2.7.** [Lin99, lemma 1.7 (ii)] *Let  $f_i : B_2(p_i) \subseteq M_i \rightarrow N$  be a sequence of stationary harmonic maps satisfying (1.16) with  $\int_{B_2(p)} |\nabla f_i|^2 \leq \Lambda$ . Assume  $K_{M_i} < \delta_i \rightarrow 0$  and that  $|\theta_{i,2}(p_i) - \theta_{i,\delta_i}| < \delta_i \rightarrow 0$ . Then the defect measure  $|\nabla f_i|^2 dv_g \xrightarrow{*} |\nabla f|^2 dv_g + \nu$  is 0-symmetric, that is, both the function  $f$  and  $\nu$  are invariant under dilation around the origin.*

**2.4.  $\epsilon$ -regularity for Minimizers.** In this subsection we quickly review the  $\epsilon$ -regularity theorems of [CN13b], which themselves build on the difficult work of [SU82]. This will be our primary technical tool in upgrading the structural results on stationary harmonic maps to the regularity results for minimizing harmonic maps. Recall the definition of the regularity scale given in Definition 1.7, then the main theorem of this subsection is the following:

**Theorem 2.8** (Minimizing  $\epsilon$ -Regularity [CN13b]). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a minimizing harmonic map satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Then there exists  $\epsilon(n, K_N, \Lambda) > 0$  such that if  $K_M < \epsilon$  and  $B_2(p)$  is  $(n - 2, \epsilon)$ -symmetric, then*

$$r_f(p) \geq 1.$$

*Remark 2.4.* As previously remarked, the assumption  $K_M < \epsilon$  is of little consequence, since this just means focusing the estimates on balls of sufficiently small radius after rescaling.

The following stationary version was proved in [CHN13a], and is essentially just a combination of the defect measure ideas of [Lin99], the  $\epsilon$ -regularity of [Bet93], and a contradiction argument:

**Theorem 2.9** (Stationary  $\epsilon$ -Regularity [CHN13a]). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a minimizing harmonic map satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Assume that there exists no nonconstant stationary harmonic map  $S^2 \rightarrow N$ . Then there exists  $\epsilon(n, N, \Lambda) > 0$  such that if  $K_M < \epsilon$  and  $B_1(p)$  is  $(n - 2, \epsilon)$ -symmetric, then*

$$r_f(p) \geq \frac{1}{2}.$$

Let us also discuss some improvements of the above  $\epsilon$ -regularity theorems in the case where there are no nonconstant harmonic maps from  $S^k$  into  $N$ . These are the key  $\epsilon$ -regularity results needed for Theorems 1.9 and 1.10. We begin with the minimizing case:



**Theorem 2.10** (Improved Minimizing  $\epsilon$ -Regularity [CHN13a]). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a minimizing harmonic mapping satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Assume further that for some  $k \geq 2$  there exists no nonconstant minimizing harmonic maps  $S^\ell \rightarrow N$  for all  $\ell \leq k$ . Then there exists  $\epsilon(n, N, \Lambda)$  such that if  $K_M < \epsilon$  and  $B_1(p)$  is  $(n - k - 1, \epsilon)$ -symmetric, then*

$$r_f(p) \geq \frac{1}{2}.$$

Finally, we end with the stationary version of the above theorem:

**Theorem 2.11** (Improved Stationary  $\epsilon$ -Regularity [CN13b]). *Let  $f : B_2(p) \subseteq M \rightarrow N$  be a stationary harmonic mapping satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . Assume further that for some  $k \geq 2$  there exists no nonconstant stationary harmonic maps  $S^\ell \rightarrow N$  for all  $\ell \leq k$ . Then there exists  $\epsilon(n, N, \Lambda)$  such that if  $K_M < \epsilon$  and  $B_1(p)$  is  $(n - k - 1, \epsilon)$ -symmetric, then*

$$r_f(p) \geq \frac{1}{2}.$$

*Remark 2.5.* Note that in both theorems above, nothing would change if we only required that  $N$  admits no *continuous* (or even *smooth*) nonconstant stationary (or minimizing) harmonic map  $S^\ell \rightarrow N$  for all  $\ell \leq k$ . Indeed, the tangent maps at singular points of non-smooth stationary (or minimizing) harmonic maps give rise to smooth stationary (or minimizing) harmonic maps from lower dimensional spheres.

**2.5. Hausdorff, Minkowski, and packing Content.** In this subsection we give a brief review of the notions of Hausdorff, Minkowski, and packing content. We will also use this to recall the definition of Hausdorff measure. The results of this subsection are completely standard, but this gives us an opportunity to introduce some notation for the paper. For a more detailed reference, we refer the reader to [Mat95, Fed69]. Let us begin with the notions of content:

**Definition 2.12.** Given a set  $S \subseteq \mathbb{R}^n$  and  $r > 0$  we define the following:

- (1) The  $k$ -dimensional Hausdorff  $r$ -content of  $S$  is given by

$$\lambda_r^k(S) \equiv \inf \left\{ \sum \omega_k r_i^k : S \subseteq \bigcup B_{r_i}(x_i) \text{ and } r_i \leq r \right\}. \quad (2.6)$$

- (2) The  $k$ -dimensional Minkowski  $r$ -content of  $S$  is given by

$$m_r^k(S) \equiv (2r)^{s-n} \text{Vol}(B_r(S)). \quad (2.7)$$

- (3) The  $k$ -dimensional packing  $r$ -content of  $S$  is given by

$$p_r^k(S) \equiv \sup \left\{ \sum \omega_k r_i^k : x_i \in S \text{ and } \{B_{r_i}(x_i)\} \text{ are disjoint and } r_i \leq r \right\}. \quad (2.8)$$

These definitions make sense for any  $k \in [0, \infty)$ , though in this paper we will be particularly interested in integer valued  $k$ . Notice that if  $S$  is a compact set then  $\lambda_r^k(S), m_r^k(S) < \infty$  for any  $r > 0$ , and that we always have the relations

$$\lambda_r^k(S) \lesssim m_r^k(S) \lesssim p_r^k(S). \quad (2.9)$$

In particular, bounding the Hausdorff content is less powerful than bounding the Minkowski content, which is itself less powerful than bounding the packing content.

Primarily in this paper we will be mostly interested in content estimates, because these are the most effective estimates. However, since it is classical, let us go ahead and use the Hausdorff content to define a measure. To accomplish this, let us more generally observe that if  $r \leq r'$  then  $\lambda_r^k(S) \geq \lambda_{r'}^k(S)$ . In particular, we can define the limit

$$\lambda_0^k(S) \equiv \lim_{r \rightarrow 0} \lambda_r^k(S) = \sup_{r > 0} \lambda_r^k(S).$$

It turns out that  $\lambda_0^k$  is a genuine measure.

**Definition 2.13.** Given a set  $S \subseteq \mathbb{R}^n$  we define its  $k$ -dimensional Hausdorff measure by  $\lambda^k(S) \equiv \lambda_0^k(S)$ .

Similar constructions can be carried out for the Minkowski and packing content. In particular, we can define

$$\overline{m}_0^k(S) \equiv \limsup_{r \rightarrow 0} m_r^k(S), \quad \underline{m}_0^k(S) \equiv \liminf_{r \rightarrow 0} m_r^k(S), \quad (2.10)$$

$$p_0^k(S) = \lim_{r \rightarrow 0} p_r^k(S) = \inf_{r > 0} p_r^k(S). \quad (2.11)$$

**Definition 2.14.** Given a set  $S \subseteq \mathbb{R}^n$  we define its Hausdorff and Minkowski dimension (or box-dimension) by

$$\begin{aligned} \dim_H S &\equiv \inf \{k \geq 0 : \lambda_0^k(S) = 0\}, \\ \dim_M S &\equiv \inf \{k \geq 0 : \overline{m}_0^k(S) = 0\}. \end{aligned} \quad (2.12)$$

*Remark 2.6.* Note that we could define an upper and lower Minkowski dimension by

$$\overline{\dim}_M S \equiv \inf \{k \geq 0 : \overline{m}_0^k(S) = 0\}, \quad \underline{\dim}_M S \equiv \inf \{k \geq 0 : \underline{m}_0^k(S) = 0\}. \quad (2.13)$$

In general,  $\underline{\dim}_M S \leq \overline{\dim}_M S$ , where the inequality may be strict. However, for the purposes of this paper we will only be interested in the *upper* Minkowski dimension.

As an easy example consider the rationals  $\mathbb{Q}^n \subseteq \mathbb{R}^n$ . Then it is a worthwhile exercise to check that  $\dim_H \mathbb{Q}^n = 0$ , while  $\dim_M \mathbb{Q}^n = n$ .

A very important notion related to measures is the *density* at a point. Although this is standard, for completeness we briefly recall the definition of Hausdorff density, and refer the reader to [Mat95, chapter 6] for more on this subject.

**Definition 2.15.** Given a set  $S \subset \mathbb{R}^n$  which is  $\lambda^k$ -measurable, and  $x \in \mathbb{R}^n$ , we define the  $k$ -dimensional upper and lower density of  $S$  at  $x$  by

$$\theta^{*k}(S, x) = \limsup_{r \rightarrow 0} \frac{\lambda^k(S \cap B_r(x))}{\omega_k r^k}, \quad \theta_{*}^k(S, x) = \liminf_{r \rightarrow 0} \frac{\lambda^k(S \cap B_r(x))}{\omega_k r^k}. \quad (2.14)$$

In the following, we will use the fact that for almost any point in a set with finite  $\lambda^k$ -measure, the density is bounded from above and below.

**Proposition 2.16** ([Mat95]). *Let  $S \subset \mathbb{R}^n$  be a set with  $\lambda^k(S) < \infty$ . Then for  $k$ -a.e.  $x \in S$ :*

$$2^{-k} \leq \theta^{*k}(S, x) \leq 1, \quad (2.15)$$

while for  $k$ -a.e.  $x \in \mathbb{R}^n \setminus S$

$$\theta^{*k}(S, x) = 0. \quad (2.16)$$

**2.6. The Classical Reifenberg Theorem.** In this Section we recall the classical Reifenberg Theorem, as well as some more recent generalizations. The Reifenberg theorem gives criteria on a closed subset  $S \subseteq B_2 \subseteq \mathbb{R}^n$  which determine when  $S \cap B_1$  is bi-Hölder to a ball  $B_1(0^k)$  in a smaller dimensional Euclidean space. The criteria itself is based on the existence of *good* best approximating subspaces at each scale. We start by recalling the Hausdorff distance.

**Definition 2.17.** Given two sets  $A, B \subseteq \mathbb{R}^n$ , we define the Hausdorff distance between these two by

$$d_H(A, B) = \inf \{r \geq 0 \text{ s.t. } A \subset B_r(B) \text{ and } B \subset B_r(A)\}. \quad (2.17)$$

Recall that  $d_H$  is a distance on closed sets, meaning that  $d_H(A, B) = 0$  implies  $\overline{A} = \overline{B}$ .

The classical Reifenberg theorem says the following:

**Theorem 2.18** (Reifenberg Theorem [Rei60, Sim]). *For each  $0 < \alpha < 1$  and  $\epsilon > 0$  there exists  $\delta(n, \alpha, \epsilon) > 0$  such that the following holds. Assume  $0^n \in S \subseteq B_2 \subseteq \mathbb{R}^n$  is a closed subset, and that for each  $B_r(x) \subseteq B_1$  with  $x \in S$  we have*

$$\inf_{L^k} d_H(S \cap B_r(x), L^k \cap B_r(x)) < \delta r, \quad (2.18)$$

where the inf is taken over all  $k$ -dimensional affine subspaces  $L^k \subseteq \mathbb{R}^n$ . Then there exists  $\phi : B_1(0^k) \rightarrow S$  which is a  $C^\alpha$  bi-Hölder homeomorphism onto its image with  $[\phi]_{C^\alpha}, [\phi^{-1}]_{C^\alpha} < 1 + \epsilon$  and  $S \cap B_1 \subseteq \phi(B_1(0^k))$ .

*Remark 2.7.* In fact, one can prove a little more. In particular, under the hypothesis of the previous theorem, there exists a closed subset  $S' \subset \mathbb{R}^n$  such that  $S' \cap B_1(0) = S \cap B_1(0)$  and which is homeomorphic to a  $k$ -dimensional subspace  $0^n \in T_0 \subseteq \mathbb{R}^n$  via the  $C^\alpha$  bi-Hölder homeomorphism  $\phi : T_0 \rightarrow S'$ . Moreover,  $|\phi(x) - x| \leq C(n)\delta$  for all  $x \in T_0$  and  $\phi(x) = x$  for all  $x \in T_0 \setminus B_2(0)$ .

One can paraphrase the above to say that if  $S$  can be well approximated on every ball by a subspace in the  $L^\infty$ -sense, then  $S$  must be bi-Hölder to a ball in Euclidean space.

Let us also mention that there are several more recent generalizations of the classic Reifenberg theorem. In [Tor95], the author proves a strengthened version of (2.18) that allows one to improve bi-Hölder to bi-Lipschitz. Unfortunately, for the applications of this paper the hypotheses of [Tor95] are much too restrictive. We will require a weaker condition than in [Tor95], which is more integral in nature, see Theorem 3.2. In exchange, we will improve the bi-Hölder of the classical Reifenberg to  $W^{1,p}$ .

We will also need a version of the classical Reifenberg which only assumes that the subset  $S$  is contained near a subspace, not conversely that the subspace is also contained near  $S$ . In exchange, we will only conclude the set is rectifiable. A result in this direction was first proved in [DT12], but again the hypotheses are too restrictive for the applications of this paper, and additionally there is a topological assumption necessary for the results of [DT12], which is not reasonable in the context in this paper. We will see how to appropriately drop this assumption in Theorem 3.3.

**2.7.  $W^{1,p}$  maps and rectifiability.** In the paper we will be using the structure of  $W^{1,p}$  maps for  $p > k$  in order to conclude rectifiable structures on sets. For the reader's convenience, here we recall a standard result about rectifiability and  $W^{1,p}$  maps:

**Lemma 2.19.** *Let  $\Omega \subset \mathbb{R}^k$  be an open domain, and let  $f : \Omega \rightarrow \mathbb{R}^n$  be a  $W^{1,p}$  map with  $p > k$ . Then for all  $K \subset \Omega$ ,  $f(K)$  is a  $k$ -rectifiable set.*

In order to prove this we need, by definition, to show that there exists a countable sequence of Lipschitz maps  $f_i : \mathbb{R}^k \rightarrow \mathbb{R}^n$  such that

$$\lambda^k \left( f(K) \setminus \bigcup_i f_i(\mathbb{R}^k) \right) = 0. \quad (2.19)$$

By a classical result, it is possible for  $p > k$  to approximate all  $W^{1,p}$  maps with Lipschitz maps. More specifically, for all  $\epsilon > 0$ , there exists a Lipschitz map  $f_\epsilon$  which coincides with  $f$  up to a set  $E_\epsilon$  of small  $k$ -dimensional measure in  $\mathbb{R}^k$ . In particular, we have

$$E_\epsilon \equiv \left\{ x \in \mathbb{R}^k \text{ s.t. } f(x) \neq f_\epsilon(x) \right\}, \quad \lambda^k(E_\epsilon) < \epsilon. \quad (2.20)$$

For a detailed reference, see for example [EG92, theorem 3, sec 6.6.3] or [Zie89, section 3.10]). The only thing left to prove is that  $f(E_\epsilon)$  has small measure:

**Lemma 2.20.** *Let  $f$  be the continuous representative of a  $W^{1,p}$  map as above with  $p > k$ . Then for all measurable subsets  $E \subset \Omega$  we have*

$$\lambda^k(f(E)) \leq C(k, n, p) \|\nabla f\|_{L^p(\Omega)}^k \left( \lambda^k(E) \right)^{1-\frac{k}{p}}. \quad (2.21)$$

This lemma follows from standard Morrey-type estimates, see for example [BMT13, proposition 2.4].

### 3. THE $W^{1,p}$ -REIFENBERG AND RECTIFIABLE-REIFENBERG THEOREMS

In this Section we state very general Reifenberg type theorems for subsets of Euclidean spaces, their proofs are carried out in Sections 5 and 6. Although this is a key ingredient of the proof of our main theorems, the results of this Section are of an independent nature, and may be of some separate interest. Therefore, we will attempt to write them so that this Section may be read independently of the rest of the paper, with the exception of the appropriate Preliminary Sections.

As described in Section 2.6, the classical Reifenberg theorem tells us that if a closed subset  $S \subseteq B_2 \subseteq \mathbb{R}^n$  can be well approximated on every ball by some  $k$ -dimensional affine space in the  $L^\infty$  sense, then  $S \cap B_1$  must itself be homeomorphic, in fact bi-Hölder, to a ball  $B_1(0^k)$  in a smaller dimensional Euclidean space.

For the applications of this paper, we will need to improve on the classical Reifenberg theorem in several ways. At the most basic level,  $C^\alpha$ -equivalence is not strong enough. We will require control on the gradient of the mappings, since this is the essence of rectifiability and volume control. For this purpose we will obtain  $W^{1,p}$ -control on our mappings for  $k < p < \infty$ . We will deal with this first in Theorem 3.2, where we will also replace the  $L^\infty$  closeness of the Reifenberg theorem with an  $L^2$  closeness condition, which turns out to be more natural in the applications.

A second manner in which we will need improvement is that we will need to allow for the existence of holes. That is, in applications our set  $S$  may be rectifiable, but it might not be homeomorphic to a ball, this is much too strong. To deal with this we can weaken the hypothesis of the Reifenberg theorem and only assume that on each ball  $B_r(x)$  there is a  $k$ -dimensional subspace  $V$  such that each point of our set  $S$  is close to  $V$ , but not conversely that each point of the subspace  $V$  is close to  $S$ . We study this case in Theorem 3.3, where we will show under natural conditions that such a set  $S$  is rectifiable with volume bounds, which is easily seen to be a sharp result under the assumptions of Theorem 3.3.

The last version of the Reifenberg theorem that we prove has a more discrete nature, and it is the one we will use to obtain the volume bounds in the rest of the paper. In this case, our set  $S = \{x_i\}$  will be a discrete set of points paired with radii  $r_i > 0$  such that  $\{B_{r_i}(x_i)\}$  are disjoint balls. We will associate to this collection the measure  $\mu = \sum \omega_k r_i^k \delta_{x_i}$ , and under the appropriate assumptions show in Theorem 3.4 that the volume of  $\mu$  continues to enjoy the packing upper bound  $\mu(B_r(x)) < Cr^k$  for all radii  $r > r_i$ . In the applications, we will prove the main theorems of this paper by building a series of covers of the quantitative stratifications  $S_\epsilon^k$ , and we will apply the discrete  $W^{1,p}$ -Reifenberg to each of this to obtain volume bounds. Only at the last stage will we apply Theorem 3.3 in order to obtain the rectifiability of the sets as well.

It is worth mentioning that some interesting generalizations of Reifenberg's theorem have been proved in numerous sources. Generalizations in the spirit of this paper are explored in [DS93], [Tor95], and [DT12]. However, in each of these cases the requirements of these theorems are too stringent to be applicable in our situation. Very recently, and using techniques independent from this work, the authors in [AT15],[To15] proved necessary and sufficient conditions for the  $k$ -rectifiability of a set which are closely related to the results of this paper. However, these results lack any *a priori* control over the volume of the sets in question, which is fundamental in this paper to the applications.

**3.1. Statement of Main  $W^{1,p}$ -Reifenberg and rectifiable-Reifenberg Results.** There are three main results we wish to discuss in this subsection, each more general than the last. In order to keep the statements as clean and intuitive as possible we will introduce the following definitions, and discuss them briefly.

**Definition 3.1.** Let  $\mu$  be a measure on  $B_2$  with  $r > 0$  and  $k \in \mathbb{N}$ . Then we define the  $k$ -dimensional displacement by

$$D_\mu^k(x, r) \equiv \inf_{L^k \subseteq \mathbb{R}^n} r^{-(k+2)} \int_{B_r(x)} d^2(y, L^k) d\mu(y), \quad (3.1)$$

if  $\mu(B_r(x)) \geq \epsilon_n r^k \equiv (1000n)^{-7n^2} r^k$ , and  $D_\mu^k(x, r) \equiv 0$  otherwise, where the inf's are taken over all  $k$ -dimensional affine subspaces  $L^k \subseteq \mathbb{R}^n$ . If  $S \subseteq B_2$ , then we can define its  $k$ -displacement  $D_S^k(x, r)$  by associating to  $S$  the  $k$ -dimensional Hausdorff measure  $\lambda_S^k$  restricted to  $S$ .

*Remark 3.1.* Note that in literature  $D$  is usually referred to as the Jones' number  $\beta_2$ .

*Remark 3.2.* One can replace  $\epsilon_n$  by any smaller lower bound, zero included, and the proofs and statements will all continue to hold. Our particular choice is based on constants which will be obtained in Section 4.

*Remark 3.3.* Notice that the definitions are scale invariant. In particular, if we rescale  $B_r \rightarrow B_1$  and let  $\tilde{S}$  be the induced set, then  $D_S^k(x, r) \rightarrow D_{\tilde{S}}^k(x, 1)$ .

*Remark 3.4.* Notice the monotonicity given by the following: If  $S' \subseteq S$ , then  $D_{S'}^k(x, r) \leq D_S^k(x, r)$ .

*Remark 3.5.* It is easily seen from the definition that  $D_\mu^k(x, r)$  is controlled above and below by  $D_\mu^k(x, r/2)$  and  $D_\mu^k(x, 2r)$ , as long as  $\mu$  is not too small on  $B_r(x)$ . In particular, if  $\mu(B_r(x)) \geq 2^k \epsilon_n r^k$ , then for all  $y \in S \cap B_r(x)$ ,  $D_\mu^k(x, r) \leq 2^{k+2} D_\mu^k(y, 2r)$ . As a corollary we have the estimate

$$\mu(B_r(x)) \geq 2^k \epsilon_n r^k \implies D_\mu^k(x, r) \leq 2^{k+2} \int_{B_r(x)} D_\mu^k(y, 2r) d\mu(y). \quad (3.2)$$

The first of the main results of this part of the paper is not strictly used elsewhere in the paper, but it is a natural generalization of the Reifenberg and gives intuition and motivation for the rest of the statements, which are essentially more complicated versions of it. Thus, we have decided it is worth discussing independently. The first theorem of this section is the following:

**Theorem 3.2** ( $W^{1,p}$ -Reifenberg). *For each  $\epsilon > 0$  and  $p \in [1, \infty)$  there exists  $\delta(n, \epsilon, p) > 0$  such that the following holds. Let  $S \subseteq B_2 \subseteq \mathbb{R}^n$  be a closed subset with  $0^n \in S$ , and assume for each  $x \in S \cap B_1$  and  $B_r(x) \subseteq B_2$  that*

$$\inf_{L^k} d_H(S \cap B_r(x), L^k \cap B_r(x)) < \delta r, \quad (3.3)$$

$$\int_{S \cap B_r(x)} \left( \int_0^r D_S^k(y, s) \frac{ds}{s} \right) d\lambda^k(y) < \delta^2 r^k. \quad (3.4)$$

Then the following hold:

- (1) *there exists a mapping  $\phi : \mathbb{R}^k \rightarrow \mathbb{R}^n$  which is a  $1 + \epsilon$  bi- $W^{1,p}$  map onto its image and such that  $S \cap B_1(0^n) = \phi(B_1(0^k)) \cap B_1(0^n)$ . In particular*

$$\int_{B_1(0^k)} |\nabla \phi|^p \leq 1 + \epsilon, \quad \int_{S \cap B_1(0^n)} |\nabla \phi^{-1}|^p \leq 1 + \epsilon. \quad (3.5)$$

- (2)  *$S \cap B_1(0^n)$  is rectifiable.*

(3) For each ball  $B_r(x) \subseteq B_1$  with  $x \in S$  we have

$$(1 - \epsilon)\omega_k r^k \leq \lambda^k(S \cap B_r(x)) \leq (1 + \epsilon)\omega_k r^k. \quad (3.6)$$

*Remark 3.6.* Results (2) and (3) both follow from (1). Indeed by a standard result, if  $\phi$  is a  $W^{1,p}$  map with  $p > k$ , then its image is a  $k$ -rectifiable set (see Lemma 2.19). We get (3) by applying the result of (1) to all smaller balls  $B_r(x) \subseteq B_1$ , since the assumptions of the theorem hold on these balls as well.

Note that the  $W^{1,p}$  estimates for  $\phi^{-1}$  are justified in Lemma 6.2.

*Remark 3.7.* Note that for  $p > k$ , by Sobolev embeddings a  $W^{1,p}$  map is also a  $C^\alpha$  map with  $\alpha = 1 - \frac{k}{p}$ .

*Remark 3.8.* As it is easily seen, the requirement that  $S$  is closed is essential for this theorem, and in particular for the lower bound on the Hausdorff measure. As an example, consider any set  $S \subseteq \mathbb{R}^k$  which is dense but has zero Hausdorff measure. In the following theorems, we will not be concerned with lower bounds on the measure, and we will be able to drop the closed assumption.

Let us now consider the case when we drop the assumption (3.3) from the result. The key distinction now is that  $S$  only needs to be locally near a piece of a  $k$ -dimensional subspace, but not conversely. Thus, we cannot hope to obtain topological information about the set  $S$ , see Section 11.3. The precise statement is the following:

**Theorem 3.3** (Rectifiable-Reifenberg). *For every  $\epsilon > 0$ , there exists  $\delta(n, \epsilon) > 0$  such that the following holds. Let  $S \subseteq B_2 \subseteq \mathbb{R}^n$  be a  $\lambda^k$ -measurable subset, and assume for each  $B_r(x) \subseteq B_2$  with  $\lambda^k(S \cap B_r(x)) \geq \epsilon_n r^k$  that*

$$\int_{S \cap B_r(x)} \left( \int_0^r D_S^k(y, s) \frac{ds}{s} \right) d\lambda^k(y) < \delta^2 r^k. \quad (3.7)$$

Then the following hold:

(1) For each ball  $B_r(x) \subseteq B_1$  with  $x \in S$  we have

$$\lambda^k(S \cap B_r(x)) \leq (1 + \epsilon)\omega_k r^k. \quad (3.8)$$

(2)  $S \cap B_1(0^n)$  is  $k$ -rectifiable.

*Remark 3.9.* Notice that for the statement of the theorem we do not need control over balls which already have small measure. This will be quite convenient for the applications.

*Remark 3.10.* Instead of (3.7) we may assume the essentially equivalent estimate

$$\sum_{r_\alpha \leq 4r} \int_{S \cap B_r(x)} D_S^k(y, r_\alpha) d\lambda^k(y) < \delta^2 r^k. \quad (3.9)$$

In the applications, this will be the more convenient phrasing.

Finally, we end by stating a version of the above theorem which is more discrete in nature. This result will be particularly important in the proof of the main theorems of this paper:

**Theorem 3.4** (Discrete Rectifiable-Reifenberg). *There exists  $\delta(n) > 0$  and  $D(n)$  such that the following holds. Let  $\{B_{r_j}(x_j)\}_{x_j \in S} \subseteq B_2$  be a collection of disjoint balls with  $x_s \in B_1(0)$ , and let  $\mu \equiv \sum_{x_j \in S} \omega_k r_j^k \delta_{x_j}$  be the associated measure. Assume that for each  $B_r(x) \subseteq B_2$  with  $\mu(B_r(x)) \geq \epsilon_n r^k$  we have*

$$\int_{B_r(x)} \left( \int_0^r D_\mu^k(y, s) \frac{ds}{s} \right) d\mu(y) < \delta^2 r^k. \quad (3.10)$$

Then we have the estimate

$$\sum_{x_j \in B_1} r_j^k < D(n). \quad (3.11)$$

*Remark 3.11.* As in Theorem 3.3, instead of (3.10) we may assume the estimate

$$\sum_{r_\alpha \leq 2r} \int_{B_r(x)} D_\mu^k(y, r_\alpha) d\mu(y) < \delta^2 r^k. \quad (3.12)$$

In the applications, this will be the more convenient phrasing.

For many of the applications of this paper, it is this version of the Reifenberg which will be most important. The reasoning is that to keep uniform control on all estimates our inductive covering will need to cover at all scales. It is only at the last scale that we begin to cover the singular sets  $S^k$  on sets of positive measure.

#### 4. TECHNICAL CONSTRUCTIONS TOWARD NEW REIFENBERG RESULTS

In this section, we prove some technical lemmas needed for dealing with the relation between best  $L^2$  subspaces. These elementary results will be used in many of the estimates of subsequent sections.

**4.1. Hausdorff distance and subspaces.** We start by recalling some standard facts about affine subspaces in  $\mathbb{R}^n$  and Hausdorff distance.

**Definition 4.1.** Given two linear subspaces  $L, V \subseteq \mathbb{R}^n$ , we define the Grassmannian distance between these two as

$$d_G(L, V) = d_H(L \cap B_1(0), V \cap B_1(0)) = d_H\left(\overline{L \cap B_1(0)}, \overline{V \cap B_1(0)}\right), \quad (4.1)$$

Note that if  $\dim(L) \neq \dim(V)$ , then  $d_G(L, V) = 1$ .

For general subsets in  $\mathbb{R}^n$ , it is evident that  $A \subseteq B_\delta(B)$  does not imply  $B \subseteq B_{c\delta}(A)$ . However, if  $A$  and  $B$  are affine spaces with the same dimension, then it is not difficult to see that this property holds. More precisely:

**Lemma 4.2.** *Let  $V, W$  be two  $k$ -dimensional affine subspaces in  $\mathbb{R}^n$ , and suppose that  $V \cap B_{1/2}(0) \neq \emptyset$ . There exists a constant  $c(k, n)$  such that if  $V \cap B_1(0) \subseteq B_\delta(W \cap B_1(0))$ , then  $W \cap B_1(0) \subseteq B_{c\delta}(V \cap B_1(0))$ . Thus in particular  $d_H(V \cap B_1(0), W \cap B_1(0)) \leq c\delta$ .*



*Proof.* The proof relies on the fact that  $V$  and  $W$  have the same dimension. Let  $x_0 \in V$  be the point of minimal distance from the origin. By assumption, we have that  $\|x_0\| \leq 1/2$ . Let  $x_1, \dots, x_k \in V \cap \overline{B_1(0)}$  be a sequence of points such that

$$\|x_i - x_0\| \geq 1/2 \quad \text{and for } i \neq j, \quad \langle x_i - x_0, x_j - x_0 \rangle = 0. \quad (4.2)$$

In other words,  $\{x_i - x_0\}_{i=1}^k$  is an affine base for  $V$ . Let  $\{y_i\}_{i=0}^k \subseteq W \cap \overline{B_1(0)}$  be such that  $d(x_i, y_i) \leq \delta$ . Then

$$\|y_i - y_0\| \geq 1/2 - 2\delta \quad \text{and for } i \neq j, \quad \left| \langle y_i - y_0, y_j - y_0 \rangle \right| \leq 2\delta + 4\delta^2. \quad (4.3)$$

This implies that for  $\delta \leq \delta_0(n)$ ,  $\{y_i - y_0\}_{i=1}^k$  is an affine base for  $W$  and for all  $y \in W$

$$y = y_0 + \sum_{i=1}^k \alpha_i (y_i - y_0), \quad |\alpha_i| \leq 10 \|y - y_0\|. \quad (4.4)$$

Now let  $y \in W \cap \overline{B_1(0)}$  be the point of maximum distance from  $V$ , and let  $\pi$  be the projection onto  $V$  and  $\pi^\perp$  the projection onto  $V^\perp$ . Then

$$d(y, V) = d(y, \pi(y)) = \|\pi^\perp(y)\| \leq \|\pi^\perp(y_0 - x_0)\| + \sum_{i=1}^k |\alpha_i| \|\pi^\perp(y_i - y_0)\| \leq c'(n, k)\delta. \quad (4.5)$$

Since  $y \in \overline{B_1(0)}$ , then  $\pi(y) \in V \cap B_{1+c'\delta}(0)$ , and thus  $d(y, V \cap B_1(0)) \leq 2c'\delta \equiv c\delta$ , which proves the claim.  $\square$

Next we will see that the Grassmannian distance between two subspaces is enough to control the projections with respect to these planes. In order to do so, we recall a standard estimate.

**Lemma 4.3.** *Let  $V, W$  be linear subspaces of a Hilbert space. Then  $d_G(V, W) = d_G(V^\perp, W^\perp)$ .*

*Proof.* We will prove that  $d_G(V^\perp, W^\perp) \leq d_G(V, W)$ . By symmetry, this is sufficient.

Take  $x \in V^\perp$  such that  $\|x\| = 1$ , and consider that  $d(x, W^\perp) = \|\pi_{W^\perp}(x)\|$ . Let  $z = \pi_W(x)$  and  $y = \pi_V(z)$ . We want to show that if  $d_G(V, W) \leq \epsilon < 1$ , then  $\|z\| \leq \epsilon$ . We can limit our study to the space spanned by  $x, y, z$ , and assume wlog that  $x = (1, 0, 0)$ ,  $y = (0, b, 0)$  and  $z = (a, b, c)$ . By orthogonality between  $z$  and  $z - x$ , we have

$$a^2 + b^2 + c^2 + (1 - a)^2 + b^2 + c^2 = 1 \quad \implies \quad a = a^2 + b^2 + c^2, \quad (4.6)$$

and since  $z \in W$ , we also have  $\|z - y\| \leq \epsilon \|z\|$ , which implies

$$a^2 + c^2 \leq \epsilon^2 (a^2 + b^2 + c^2) \quad \implies \quad a^2 + c^2 \leq \frac{\epsilon^2}{1 - \epsilon^2} b^2. \quad (4.7)$$

Since the function  $f(x) = x^2/(1 - x^2)$  is monotone increasing for  $x \geq 0$ , we can define  $\alpha \geq 0$  in such a way that

$$a^2 + c^2 = \frac{\alpha^2}{1 - \alpha^2} b^2, \quad a = a^2 + b^2 + c^2 = \frac{1}{1 - \alpha^2} b^2. \quad (4.8)$$

Note that necessarily we will have  $\alpha \leq \epsilon$ . Now we have

$$\frac{1}{(1 - \alpha^2)^2} b^4 = a^2 \leq \frac{\alpha^2}{1 - \alpha^2} b^2 \quad \implies \quad b^2 \leq \alpha^2 (1 - \alpha^2) \quad \implies \quad \|z\|^2 = a^2 + b^2 + c^2 \leq \alpha^2 \leq \epsilon^2. \quad (4.9)$$

This proves that  $V^\perp \cap B_1(0) \subset B_\epsilon(W^\perp)$ . In a similar way, one proves the opposite direction.  $\square$

As a corollary, we prove that the Grassmannian distance  $d_G(V, W)$  is equivalent to the distance given by  $\|\pi_V - \pi_W\|$ .

**Lemma 4.4.** *Let  $V, W$  be linear subspaces of  $\mathbb{R}^n$ . Then for every  $x \in \mathbb{R}^n$ ,*

$$\|\pi_V(x) - \pi_W(x)\| \leq 2d_G(V, W) \|x\|. \quad (4.10)$$

*In particular, if  $x \in W^\perp$ , then  $\|\pi_V(x)\| \leq 2d_G(V, W) \|x\|$ .*

*Conversely, we have*

$$d_G(V, W) \leq \sup_{x \in \mathbb{R}^n \setminus \{0\}} \left\{ \frac{\|\pi_V(x) - \pi_W(x)\|}{\|x\|} \right\}. \quad (4.11)$$

*Proof.* The proof is just a corollary of the previous lemma. Assume wlog that  $\|x\| = 1$ , and let  $x = y + z$  where  $y = \pi_V(x)$  and  $z = \pi_{V^\perp}(x)$ . Then

$$\|\pi_V(x) - \pi_W(x)\| = \|y - \pi_W(y) - \pi_W(z)\| \leq \|y - \pi_W(y)\| + \|z - \pi_{W^\perp}(z)\| = d(y, W) + d(z, W^\perp). \quad (4.12)$$

Since  $\|y\|^2 + \|z\|^2 = \|x\|^2 = 1$ , by the previous lemma we get the first estimate.

The reverse estimate is an immediate consequence of the definition.  $\square$

**4.2. Distance between  $L^2$  best planes.** Here we study the distance between best approximating subspaces for our measure  $\mu$  on different balls. Let us begin by fixing our notation for this subsection, and pointing out the interdependencies of the constants chosen here. Throughout this subsection, our choice of scale  $\rho = \rho(n, M) > 0$  is a constant which will eventually be fixed according to Lemma 4.7. For applications to future sections, it is sufficient to know that we can take  $\rho(n, M) = 10^{-10}(100n)^{-n}M^{-1}$ . We also point out that in Section 5, we will fix  $M = 40^n$ , and so  $\rho$  will be a constant depending only on  $n$ . In particular, we can use the very coarse estimate

$$\rho = 10^{-10}(100n)^{-3n}. \quad (4.13)$$

We will also introduce a threshold value  $\gamma_k = \omega_k 40^{-k}$ . The dimensional constant  $\gamma_k$  is chosen simply to be much smaller than any covering errors which will appear.

We will consider a positive Radon measure  $\mu$  supported on  $S \subseteq B_1(0)$ , and use  $D(x, r) \equiv D_\mu^k(x, r)$  to bound the distances between best  $L^2$  planes at different points and scales. By definition let us denote by  $V(x, r)$  a best  $k$ -dimensional plane on  $B_r(x)$ , i.e., a  $k$ -dimensional affine subspaces minimizing  $\int_{B_r(x)} d(x, V)^2 d\mu$ . Note that, in general, this subspace may not be unique. We want to prove that, under reasonable hypothesis, the distance between  $V(x, r)$  and  $V(y, r')$  is small if  $d(x, y) \sim r$  and  $r' \sim r$ .

In order to achieve this, we will need to understand some minimal properties of  $\mu$ . First, we need to understand how concentrated  $\mu$  is on any given ball. For this reason, for some  $\rho > 0$  and all  $x \in B_1(0)$  we will want to consider the upper mass bound

$$\mu(B_\rho(x)) \leq M\rho^k \quad \forall x \in B_1(0). \quad (4.14)$$

However, an upper bound on the measure is not enough to guarantee best  $L^2$ -planes are close, as the following example shows:

*Example 4.1.* Let  $V, V'$  be  $k$ -dimensional subspaces,  $0 \in V \cap V'$ , and set  $S = (V \cap B_1(0) \setminus B_{1/10}(0)) \cup S'$ , where  $S' \subseteq V' \cap B_{1/10}(0)$  and  $\mu = \lambda^k|_S$ . Then evidently  $D(0, 1) \leq \lambda^k(S')$  and  $D(0, 1/10) = 0$ , independently of  $V$  and  $V'$ . However,  $V(0, 1)$  will be close to  $V$ , while  $V(0, 1/10) = V'$ . Thus, in general, we cannot expect  $V(0, 1)$  and  $V(0, 1/10)$  to be close if  $\mu(B_{1/10}(0))$  is too small.

Thus, in order to prove that the best planes are close, we need to have some definite amount of measure on the set, in such a way that  $S$  “effectively spans” a  $k$ -dimensional subspace, where by effectively span we mean the following:

**Definition 4.5.** Given a sequence of points  $p_i \in \mathbb{R}^n$ , we say that  $\{p_i\}_{i=0}^k$   $\alpha$ -effectively span a  $k$ -dimensional affine subspace if for all  $i = 1, \dots, k$

$$\|p_i - p_0\| \leq \alpha^{-1}, \quad p_i \notin B_\alpha(p_0 + \text{span}\{p_1 - p_0, \dots, p_{i-1} - p_0\}). \quad (4.15)$$

This implies that the vectors  $p_i - p_0$  are linearly independent in a quantitative way. In particular, we obtain immediately that

**Lemma 4.6.** *If  $\{p_i\}_{i=0}^k$   $\alpha$ -effectively span the  $k$ -dimensional affine subspace*

$$V = p_0 + \text{span}\{p_1 - p_0, \dots, p_k - p_0\},$$

*then for all  $x \in V$  there exists a unique set  $\{\alpha_i\}_{i=1}^k$  such that*

$$x = p_0 + \sum_{i=0}^k \alpha_i(p_i - p_0), \quad |\alpha_i| \leq c(n, \alpha) \|x - p_0\|. \quad (4.16)$$

*Proof.* The proof is quite straightforward. Since  $\{p_i - p_0\}_{i=1}^k$  are linearly independent, we can apply the Gram-Schmidt orthonormalization process to obtain an orthonormal basis  $e_1, \dots, e_k$  for the linear space  $\text{span}\{p_i - p_0\}_{i=1}^k$ . By induction, it is easy to check that for all  $i$

$$e_i = \sum_{j=1}^i \alpha'_j(p_j - p_0), \quad |\alpha'_j| \leq c(n, \alpha). \quad (4.17)$$

Now the estimate follows from the fact that for all  $x \in V$

$$x = p_0 + \sum_{i=1}^k \langle x - p_0, e_i \rangle e_i. \quad (4.18)$$

□

With these definitions, we are ready to prove that in case  $\mu$  is not too small, then its support must effectively span something  $k$ -dimensional.

**Lemma 4.7.** *Let  $\gamma_k = \omega_k 40^{-k}$ . There exists a  $\rho_0(n, \gamma_k, M) = \rho_0(n, M)$  such that if (4.14) holds for some  $\rho \leq \rho_0$  and if  $\mu(B_1(0)) \geq \gamma_k$ , then for every affine subspace  $V \subseteq \mathbb{R}^n$  of dimension  $\leq k - 1$ , there exists an  $x \in S$  such that  $B_\rho(x) \subseteq B_1(0)$ ,  $B_{10\rho}(x) \cap V = \emptyset$  and  $\mu(B_\rho(x)) \geq c(n, \rho) = c(n)\rho^n > 0$ .*

*Proof.* Let  $V$  be any  $k - 1$ -dimensional subspace, and consider the set  $B_{11\rho}(V)$ . Let  $B_i = B_\rho(x_i)$  be a sequence of balls that cover the set  $B_{11\rho}(V) \cap B_1(0)$  and such that  $B_i/2 \equiv B_{\rho/2}(x_i)$  are disjoint. If  $N$  is the number of these balls, then a standard covering argument gives

$$\begin{aligned} N\omega_n\rho^n/2^n &\leq \omega_{k-1}(1+\rho)^{k-1}\omega_{n-k+1}(12\rho)^{n-k+1} \leq 24^n\omega_{k-1}\omega_{n-k+1}\rho^{n-k+1} \\ \implies N &\leq 48^n \frac{\omega_{k-1}\omega_{n-k+1}}{\omega_n} \rho^{1-k}. \end{aligned} \quad (4.19)$$

By (4.14), the measure of the set  $B_{11\rho}(V)$  is bounded by

$$\mu(B_{11\rho}(V)) \leq \sum_i \mu(B_i) \leq MN\omega_k\rho^k = 48^n \frac{\omega_k\omega_{k-1}\omega_{n-k+1}}{\omega_n} M\rho \leq 10^5(50n)^n M\rho = c(n)M\rho. \quad (4.20)$$

Thus if

$$\rho \leq 10^{-5}(50n)^{-n}\gamma_k/(4M), \quad (4.21)$$

then  $\mu(B_{11\rho}(V)) \leq \gamma_k/4$ . In particular, we get that there must be some point of  $S$  not in  $B_{11\rho}(V)$ . More effectively, let us consider the set  $B_1(0) \setminus B_{11\rho}(V)$ . This set can be covered by at most  $c(n, \rho) = 4^n\rho^{-n}$  balls of radius  $\rho$ , and we also see that

$$\mu(B_1(0) \setminus B_{11\rho}(V)) \geq \frac{3\gamma_k}{4}. \quad (4.22)$$

Thus, there must exist at least one ball of radius  $\rho$  which is disjoint from  $B_{10\rho}(V)$  and such that

$$\mu(B_\rho(x)) \geq \frac{3\gamma_k}{4}4^{-n}\rho^n \geq c(n)\rho^n. \quad (4.23)$$

□

Now if at two consecutive scales there are some balls on which the measure  $\mu$  effectively spans  $k$ -dimensional subspaces, we show that these subspaces have to be close together.

**Lemma 4.8.** *Let  $\mu$  be a positive Radon measure and assume  $\mu(B_1(0)) \geq \gamma_k$  and that for each  $y \in B_1(0)$  we have  $\mu(B_{\rho^2}(y)) \leq M\rho^{2k}$ , where  $\rho \leq \rho_0$ . Additionally, let  $B_\rho(x) \subset B_1(0)$  be a ball such that  $\mu(B_\rho(x)) \geq \gamma_k\rho^k$ . Then if  $A = V(0, 1) \cap B_\rho(x)$  and  $B = V(x, \rho) \cap B_\rho(x)$  are  $L^2$ -best subspace approximations of  $\mu$  with  $d(x, A) < \rho/2$ , then*

$$d_H(A, B)^2 \leq c(n, \rho, M) \left( D_\mu^k(x, \rho) + D_\mu^k(0, 1) \right). \quad (4.24)$$

*Proof.* Let us begin by observing that if  $c(n, \rho, M) > 2\delta^{-1}(n, \rho, M)$ , which will be chosen later, then we may assume without loss of generality that

$$D_\mu^k(x, \rho) + D_\mu^k(0, 1) \leq \delta = \delta(n, \rho, M), \quad (4.25)$$

since otherwise (4.24) is trivially satisfied. Moreover, note that  $\gamma_k \gg \epsilon_n$ , so equation (3.1) is valid on  $B_\rho(x)$  and on  $B_1(0)$ .

We will estimate the distance  $d_H(A, B)$  by finding  $k + 1$  balls  $B_{\rho^2}(y_i)$  which have enough mass and effectively span in the appropriate sense  $V(x, \rho)$ . Given the upper bounds on  $D_\mu^k$ , we will then be in a position to prove our estimate.

Consider any  $B_{\rho^2}(y) \subseteq B_1(0)$  and let  $p(y) \in B_{\rho^2}(y)$  be the center of mass of  $\mu$  restricted to this ball. Let also  $\pi(p)$  be the orthogonal projection of  $p$  onto  $V(x, \rho)$ . By Jensen's inequality:

$$d(p(y), V(x, \rho))^2 = d(p(y), \pi(p(y)))^2 = d\left(\int_{B_{\rho^2}(y)} z d\mu(z), V(x, \rho)\right)^2 \leq \frac{1}{\mu(B_{\rho^2}(y))} \int_{B_{\rho^2}(y)} d(z, V(x, \rho))^2 d\mu(z). \quad (4.26)$$

Using this estimate and Lemma 4.7, we want to prove that there exists a sequence of  $k+1$  balls  $B_{\rho^2}(y_i) \subseteq B_\rho(x)$  such that

- (i)  $\mu(B_{\rho^2}(y_i)) \geq c(n, \rho, M) > 0$
- (ii)  $\{\pi(p(y_i))\}_{i=0}^k \equiv \{\pi_i\}_{i=0}^k$  effectively spans  $V(x, \rho)$ . In other words for all  $i = 1, \dots, k$ ,  $\pi_i \in V(x, \rho)$  and
 
$$\pi_i \notin B_{5\rho^2}(\pi_0 + \text{span}(\pi_1 - \pi_0, \dots, \pi_{i-1} - \pi_0)). \quad (4.27)$$

We prove this statement by induction on  $i = 0, \dots, k$ . For  $i = 0$ , the statement is trivially true since  $\mu(B_\rho(x)) \geq \gamma_k \rho^k$ . In order to find  $y_{i+1}$ , consider the  $i$ -dimensional affine subspace  $V^{(i)} = \text{span}(\pi_0, \dots, \pi_{i-1}) \leq V(x, \rho)$ . By Lemma 4.7 applied to the ball  $B_\rho(x)$ , there exists some  $B_{\rho^2}(y_{i+1})$  such that  $\mu(B_{\rho^2}(y_{i+1})) \geq c(n, \rho, M) > 0$  and

$$y_{i+1} \notin B_{10\rho^2}(\text{span}(\pi_0, \dots, \pi_{i-1})). \quad (4.28)$$

By definition of center of mass, it is clear that  $d(y_i, p(y_i)) \leq \rho^2$ . Moreover, by item (i) and equation (4.26), we get

$$d(p(y_{i+1}), V(x, \rho))^2 \leq c \int_{B_{\rho^2}(y)} d(z, V(x, \rho))^2 d\mu(z) \leq c D_\mu^k(x, \rho) \leq c\delta. \quad (4.29)$$

Thus by the triangle inequality we have  $d(y_i, \pi_i) \leq 2\rho^2$  if  $\delta \leq \delta_0(n, \rho, M)$  is small enough. This implies (4.27). Using similar estimates, we also prove  $d(p(y_i), V(0, 1))^2 \leq c' D_\mu^k(0, 1)$  for all  $i = 0, \dots, k$ . Thus by the triangle inequality

$$d(\pi_i, V(0, 1)) \leq d(\pi_i, p(y_i)) + d(p(y_i), V(0, 1)) \leq c(n, \rho, M) \left( D_\mu^k(x, \rho) + D_\mu^k(0, 1) \right)^{1/2}. \quad (4.30)$$

Now consider any  $y \in V(x, \rho)$ . By item (ii) and Lemma 4.6, there exists a unique set  $\{\beta_i\}_{i=1}^k$  such that

$$y = \pi_0 + \sum_{i=1}^k \beta_i (\pi_i - \pi_0), \quad |\beta_i| \leq c(n, \rho) \|y - \pi_0\|. \quad (4.31)$$

Hence for all  $y \in V(x, \rho) \cap B_\rho(x_i)$ , we have

$$d(y, V(0, 1)) \leq d(\pi_0, V(0, 1)) + \sum_i |\beta_i| d(\pi_i - \pi_0, V(0, 1)) \leq c(n, \rho, M) \left( D_\mu^k(x, \rho) + D_\mu^k(0, 1) \right)^{1/2}. \quad (4.32)$$

By Lemma 4.2, this completes the proof of (4.24).  $\square$

**4.3. Comparison between  $L^2$  and  $L^\infty$  planes.** Given  $B_r(x)$ , we denote as before by  $V(x, r)$  one of the  $k$ -dimensional subspace minimizing  $\int_{B_r(x)} d(y, V)^2 d\mu$ . Suppose that the support of  $\mu$  satisfies a uniform one-sided Reifenberg condition, i.e. suppose that there exists a  $k$ -dimensional plane  $L(x, r)$  such that  $x \in L(x, r)$  and

$$\text{supp}(\mu) \cap B_r(x) \subseteq B_{\delta r}(L(x, r)) . \quad (4.33)$$

Then, by the same technique used in Lemma 4.8, we can prove that

**Lemma 4.9.** *Let  $\mu$  be a positive Radon measure with  $\mu(B_1(0)) \geq \gamma_k$  and such that for all  $B_\rho(y) \subseteq B_1(0)$  we have  $\mu(B_\rho(y)) \leq M\rho^k$  and (4.33). Then*

$$d_H(L(0, 1) \cap B_1(0), V(0, 1) \cap B_1(0))^2 \leq c(n, \rho, M) \left( \delta^2 + D_\mu^k(0, 1) \right) . \quad (4.34)$$

**4.4. bi-Lipschitz equivalences.** In this subsection, we study a particular class of maps with nice local properties. These maps are a slightly modified version of the maps which are usually exploited to prove Reifenberg's theorem, see for example [Rei60, Tor95, DT12], [Mor66, section 10.5] or [Sim]. The estimates in this section are standard in literature.

We start by defining the functions  $\sigma$ . For some  $r > 0$ , let  $\{x_i\}$  be an  $r$ -separated subset of  $\mathbb{R}^n$ , i.e.,

$$(i) \quad d(x_i, x_j) \geq r .$$

Let also  $p_i$  be a points in  $\mathbb{R}^n$  with

$$(ii) \quad p_i \in B_{10r}(x_i)$$

and let  $V_i$  be a sequence of  $k$ -dimensional linear subspaces.

By standard theory, it is easy to find a locally finite smooth partition of unity  $\lambda_i : \mathbb{R}^n \rightarrow [0, 1]$  such that

$$(iii) \quad \text{supp}(\lambda_i) \subseteq B_{3r}(x_i) \text{ for all } i ,$$

$$(iv) \quad \text{for all } x \in \bigcup_i B_{2r}(x_i), \sum_i \lambda_i(x) = 1 ,$$

$$(v) \quad \sup_i \|\nabla \lambda_i\|_\infty \leq c(n)/r ,$$

$$(vi) \quad \text{if we set } 1 - \psi(x) = \sum_i \lambda_i(x), \text{ then } \psi \text{ is a smooth function with } \|\nabla \psi\|_\infty \leq c(n)/r .$$

Note that by (iii), and since  $x_i$  is  $r$ -separated, there exists a constant  $c(n)$  such that for all  $x$ ,  $\lambda_i(x) > 0$  for at most  $c(n)$  different indexes.

For convenience of notation, set  $\pi_V(v)$  to be the orthogonal projection onto the linear subspace  $V$  of the free vector  $v$ , and set

$$\pi_{p_i, V_i}(x) = p_i + \pi_{V_i}(x - p_i) . \quad (4.35)$$

In other words,  $\pi_{p_i, V_i}$  is the affine projection onto the affine subspace  $p_i + V_i$ . Recall that  $\pi_{V_i}$  is a linear map, and so the gradients of  $\pi_{V_i}$  and of  $\pi_{p_i, V_i}$  at every point are equal to  $\pi_{V_i}$ .

**Definition 4.10.** Given  $\{x_i, p_i, \lambda_i\}$  satisfying (i) to (vi), and given a family of linear  $k$ -dimensional spaces  $V_i$ , we define a smooth function  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^n$  by

$$\sigma(x) = x + \sum_i \lambda_i(x) \pi_{V_i^\perp}(p_i - x) = \psi(x)x + \sum_i \lambda_i(x) \pi_{p_i, V_i}(x) . \quad (4.36)$$

By local finiteness, it is evident that  $\sigma$  is smooth. Moreover, if  $\psi(x) = 1$ , then  $\sigma(x) = x$ . It is clear that philosophically  $\sigma$  is a form of “smooth interpolation” between the identity and the projections onto the subspaces  $V_i$ . It stands to reason that if  $V_i$  are all close together, then this map  $\sigma$  is close to being an orthogonal projection in the region  $\bigcup_i B_{2r}(x_i)$ .

**Lemma 4.11.** *Suppose that there exists a  $k$ -dimensional linear subspace  $V \subseteq \mathbb{R}^n$  and a point  $p \in \mathbb{R}^n$  such that for all  $i$*

$$d_G(V_i, V) \leq \delta, \quad d(p_i, p + V) \leq \delta. \quad (4.37)$$

Then the map  $\sigma$  restricted to the set  $U = \psi^{-1}(0) = (\sum_i \lambda_i)^{-1}(1)$  can be written as

$$\sigma(x) = \pi_{p,V}(x) + e(x), \quad (4.38)$$

and  $e(x)$  is a smooth function with

$$\|e\|_\infty + \|\nabla e\|_\infty \leq c(n)\delta/r = c(n, r)\delta. \quad (4.39)$$

*Remark 4.1.* Thus, on  $U$  we have that  $\sigma$  is the affine projection onto  $V$  plus an error which is small in  $C^1$ .

*Proof.* On the set  $U$ , we can define

$$\begin{aligned} e(x) &= \sigma(x) - \pi_{p,V}(x) = -\pi_{p,V}(x) + \sum_i \lambda_i(x) \cdot (\pi_{p_i, V_i}(x)) = \\ &= \sum_i \lambda_i(x) \cdot (p_i - p - \pi_V(p_i - p) + \pi_V(p_i) - \pi_{V_i}(p_i) + \pi_{V_i}(x) - \pi_V(x)). \end{aligned} \quad (4.40)$$

By (4.37) and Lemma 4.4, we have the estimates

$$\|p_i - p - \pi_V(p_i - p)\| < \delta, \quad \|\pi_V(x - p_i) - \pi_{V_i}(x - p_i)\| \leq c(n)\delta \|x - p_i\| \leq 13c(n)r\delta. \quad (4.41)$$

This implies

$$\|e\|_{L^\infty(U)} \leq c(n)(1 + 13r)\delta \leq c(n)\delta. \quad (4.42)$$

As for  $\nabla e$ , we have

$$\nabla e = \sum_i \nabla \lambda_i(x) \cdot (p_i - p - \pi_V(p_i - p) + \pi_V(p_i) - \pi_{V_i}(p_i) + \pi_{V_i}(x) - \pi_V(x)) + \sum_i \lambda_i(x) \nabla (\pi_{V_i}(x) - \pi_V(x)). \quad (4.43)$$

The first sum is easily estimated, and since  $\langle \nabla(\pi_W)|_x, w \rangle = \pi_W(w)$ , we can still apply Lemma 4.4 and conclude:

$$\|\nabla e\|_{L^\infty(U)} \leq \frac{c(n)}{r} \delta. \quad (4.44)$$

□

As we have seen,  $\sigma$  is in some sense close to the affine projection to  $p + V$ . In the next lemma, which is similar in spirit to [Sim, squash lemma], we prove that the image through  $\sigma$  of a graph over  $V$  is again a graph over  $V$  with nice bounds.

**Lemma 4.12** (squash lemma). *Fix  $\rho \leq 1$  and some  $B_{r/\rho}(y) \subseteq \mathbb{R}^n$ , let  $I = \{x_i\} \cap B_{5r/\rho}(y)$  be an  $r$ -separated set and define  $\sigma$  as in Definition 4.10. Suppose that there exists a  $k$ -dimensional subspace  $V$  and some  $p \in \mathbb{R}^n$  such that  $d(y, p + V) \leq \delta r$  and for all  $i$ :*

$$d(p_i, p + V) \leq \delta r \quad \text{and} \quad d_G(V_i, V) \leq \delta. \quad (4.45)$$

Suppose also that there exists a  $C^1$  function  $g : V \rightarrow V^\perp$  such that  $G \subseteq \mathbb{R}^n$  is the graph

$$G = \{p + x + g(x) \text{ for } x \in V\} \cap B_{r/\rho}(y),$$

and  $r^{-1} \|g\|_\infty + \|\nabla g\|_\infty \leq \delta'$ . There exists a  $\delta_0(n) > 0$  sufficiently small such that if  $\delta \leq \delta_0 \rho$  and  $\delta' \leq 1$ , then

- (i)  $\forall z \in G$ ,  $r^{-1} |\sigma(z) - z| \leq c(n)(\delta + \delta')\rho^{-1}$ , and  $\sigma$  is a  $C^1$  diffeomorphism from  $G$  to its image,
- (ii) the set  $\sigma(G)$  is contained in a  $C^1$  graph  $\{p + x + \tilde{g}(x), x \in V\}$  with

$$r^{-1} \|\tilde{g}\|_\infty + \|\nabla \tilde{g}\|_\infty \leq c(n)(\delta + \delta')\rho^{-1}. \quad (4.46)$$

- (iii) moreover, if  $U'$  is such that  $B_{c(\delta + \delta')\rho^{-1}}(U') \subseteq \psi^{-1}(0)$ , then the previous bound is independent of  $\delta'$ , in the sense that

$$r^{-1} \|\tilde{g}\|_{L^\infty(U' \cap V)} + \|\nabla \tilde{g}\|_{L^\infty(U' \cap V)} \leq c(n)\delta\rho^{-1}. \quad (4.47)$$

For example, if  $\delta' \leq \delta_0(n)\rho^{-1}$ , we can take  $U' = \bigcup_i B_{1.5r}(x_i)$ .

- (iv) the map  $\sigma$  is a bi-Lipschitz equivalence between  $G$  and  $\sigma(G)$  with bi-Lipschitz constant  $\leq 1 + c(n)(\delta + \delta')^2\rho^{-2}$ .

*Proof.* For convenience, we fix  $r = 1$  and  $p = 0$ . By notation, given any map  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $p \in \mathbb{R}^n$  and  $w \in T_p(\mathbb{R}^n) = \mathbb{R}^n$ , we will denote by  $\nabla|_p f[w]$  the gradient of  $f$  evaluated at  $p$  and applied to the vector  $w$ .

Recall that

$$\sigma(x + g(x)) = \psi(z)(x + g(x)) + \sum_{x_i \in I} \lambda_i(z) \left( \pi_{p_i, V_i}(x + g(x)) \right), \quad 1 - \psi(x) = \sum_{x_i \in I} \lambda_i(x), \quad (4.48)$$

where we have set for convenience  $z = z(x) = x + g(x)$ . Define  $h(x)$  by

$$(1 - \psi(z))x + h(x) \equiv \sum_i \lambda_i(z) \left( \pi_{p_i, V_i}(x + g(x)) \right). \quad (4.49)$$

Set also  $h^T(x) = \pi_V(h(x))$  and  $h^\perp(x) = \pi_{V^\perp}(h(x))$ . By projecting the function  $\sigma(x + g(x))$  onto  $V$  and its orthogonal complement we obtain

$$\begin{aligned} \sigma(x + g(x)) &\equiv \sigma^T(x) + \sigma^\perp(x), \\ \sigma^T(x) &= x + h^T(x), \quad \sigma^\perp(x) = \psi(z)g(x) + h^\perp(x). \end{aligned} \quad (4.50)$$

We claim that if  $\delta' \leq 1$ , then

$$|h^T(x)| + |\nabla h^T(x)| \leq \frac{c\delta}{\rho}, \quad (4.51)$$

where this bound is independent of  $\delta'$  as long as  $\delta' \leq 1$ . Indeed, for all  $x \in V$  we have

$$h^T(x) = \pi_V \left[ \sum_i \lambda_i(z) \left( \pi_{p_i, V_i}(x + g(x)) - x \right) \right] = \sum_i \lambda_i(z) \pi_V \left[ \left( \pi_{p_i, V_i}(x) - \pi_V(x) \right) + \pi_{p_i, V_i}(g(x)) \right] \quad (4.52)$$



Given (4.45) and Lemma 4.4, with computations similar to (4.41), we get  $|h^T(x)| \leq c\delta(1 + \rho^{-1}) \leq c\delta\rho^{-1}$ . As for the gradient, we get for any vector  $w \in V$

$$\nabla h^T|_x[w] = \pi_V \left[ \sum_i \nabla \lambda_i|_z [w + \nabla g|_x[w]] (\pi_{p_i, V_i}(x + g(x)) - x) + \sum_i \lambda_i(z) (\pi_{V_i}(w + \nabla g[w]) - w) \right], \quad (4.53)$$

In particular, we obtain

$$|\nabla h^T|_x[w]| \leq \sum_i |\nabla \lambda_i| (1 + |\nabla g|) |w| |\pi_{p_i, V_i}(x + g(x)) - x| + \sum_i \lambda_i(z) (|\pi_{V_i}(w) - w| + |\pi_{V_i}(\nabla g[w])|). \quad (4.54)$$

For the first term, we can estimate

$$|\nabla \lambda_i| \leq c(n), \quad |\nabla g| \leq \delta' \leq 1, \quad |\pi_{p_i, V_i}(x + g(x)) - x| \leq |\pi_{p_i, V_i}(x) - x| + |\pi_{V_i}(g(x))|. \quad (4.55)$$

Since  $x \in V$  with  $|x| \leq \rho^{-1}$ , and  $g(x) \in V^\perp$ , by (4.45) and Lemma 4.4 we obtain

$$|\pi_{p_i, V_i}(x + g(x)) - x| \leq c\delta\rho^{-1}. \quad (4.56)$$

As for the second term, we have

$$\|\pi_{V_i}(w) - w\| \leq c\delta |w|, \quad |\pi_{V_i}(\nabla g[w])| \leq c\delta\delta' |w| \leq c\delta |w|. \quad (4.57)$$

Summing all the contributions, we obtain (4.51) as wanted.

Thus we can apply the inverse function theorem on the function  $\sigma^T(x) : V \rightarrow V$  and obtain a  $C^1$  inverse  $Q$  such that for all  $x \in V$ ,  $|Q(x) - x| + |\nabla Q - id| \leq c(n)\delta\rho^{-1}$ , and if  $\psi(x + g(x)) = 1$ , then  $Q(x) = x$ . So we can write that for all  $x \in V$

$$\sigma(x + g(x)) = \sigma^T(x) + \tilde{g}(\sigma^T(x)) \quad \text{where} \quad \tilde{g}(x) = \sigma^\perp(Q(x)) = h^\perp(Q(x)) + \psi(z(Q(x)))g(Q(x)). \quad (4.58)$$

Arguing as above, we see that  $h^\perp(x)$  is a  $C^1$  function with

$$|h^\perp(x)| + |\nabla h^\perp(x)| \leq \frac{c\delta}{\rho}, \quad (4.59)$$

and this bound is independent of  $\delta'$  (as long as  $\delta' \leq 1$ ).

Thus the function  $\tilde{g} : V \rightarrow V^\perp$  satisfies for all  $x$  in its domain

$$|\tilde{g}(x)| + |\nabla \tilde{g}(x)| \leq c(n)(\delta + \delta')\rho^{-1}, \quad (4.60)$$

Moreover, for those  $x$  such that  $\psi(Q(x) + g(Q(x))) = 0$ , the estimates on  $\tilde{g}$  are independent of  $\delta'$ , in the sense that  $|\tilde{g}(x)| + |\nabla \tilde{g}(x)| \leq c(n)\delta\rho^{-1}$ . Note that by the previous bounds we have

$$|Q(x) + g(Q(x)) - x| \leq c(\delta + \delta')\rho^{-1}, \quad (4.61)$$

and so if  $B_{c(\delta + \delta')\rho^{-1}}(U') \subset \psi^{-1}(0)$ , then for all  $x \in U' \cap V$ ,  $\psi(Q(x) + g(Q(x))) = 0$ . This proves items (ii), (iii). As for item (i), it is an easy consequence of the estimates in (4.51), (4.59).

Now since both  $G$  and  $\sigma(G)$  are Lipschitz graphs over  $V$ , it is clear that the bi-Lipschitz map induced by  $\pi_V$  would have the right bi-Lipschitz estimate. Since  $\sigma$  is close to  $\pi_V$ , it stands to reason that this property remains true. In order to check the estimates, we need to be a bit careful about the horizontal displacement of  $\sigma$ .

**bi-Lipschitz estimates** In order to prove the estimate in (iv), we show that for all  $z = x + g(x) \in G$  and for all unit vectors  $w \in T_z(G) \subset \mathbb{R}^n$ , we have

$$\left| |\nabla \sigma|_z[w]|^2 - 1 \right| \leq c(\delta + \delta')^2. \quad (4.62)$$

First of all, note that if  $\psi(z) = 1$ , then  $\sigma$  is the identity, and there's nothing to prove.

In general, we have that

$$\nabla \sigma|_z[w] = \left( \psi(z)w + \sum_i \lambda_i(z) \pi_{V_i}[w] \right) + \left( z \nabla \psi[w] + \sum_i \pi_{p_i, V_i}(z) \nabla \lambda_i[w] \right) \equiv A + B. \quad (4.63)$$

Since  $\psi(z) + \sum_i \lambda_i(z) = 1$  everywhere by definition, we have

$$|B| = \left| \sum_i (\pi_{p_i, V_i}(z) - z) \nabla \lambda_i[w] \right| \leq c \sup_i \left\{ |\pi_{p_i, V_i}(z) - z| \right\} \leq c\delta'. \quad (4.64)$$

This last estimate comes from the fact that  $G$  is the graph of  $g$  over  $V$  with  $\|g\|_\infty \leq \delta'$ . Moreover, we can easily improve the estimate for  $B$  in the horizontal direction using Lemma 4.4. Indeed, since  $\pi_{p_i, V}(z) - z = \pi_{V^\perp}(z)$ , we have

$$|\pi_V B| = \left| \sum_i \pi_V (\pi_{p_i, V_i}(z) - z) \nabla \lambda_i[w] \right| \leq c \sup_i \left\{ \left| \pi_V (\pi_{p_i, V_i}(z) - z) \right| \right\} \leq c\delta'\delta. \quad (4.65)$$

As for  $A$ , by adapting the proof of Lemma 4.11, we get  $|A - \pi_V[w]| \leq c\delta$ . Moreover, also in this case we get better estimates for  $A$  in the horizontal direction. Indeed, we have

$$|\pi_V(A) - \pi_V[w]| = \left| \psi(z) \pi_V[w] + \sum_i (\lambda_i(z) \pi_V[\pi_{V_i}[w]]) - \pi_V[w] \right| = \left| \sum_i \lambda_i(z) (\pi_V[\pi_{V_i}[w]] - \pi_V[w]) \right|. \quad (4.66)$$

Now let  $w = \pi_V[w] + \pi_{V^\perp}[w] = w_V + w_{V^\perp}$ . Then we have

$$|\pi_V(A) - \pi_V[w]| \leq \sum_i \lambda_i(z) \left( |\pi_V[\pi_{V_i}[w_V]] - w_V| + |\pi_V[\pi_{V_i}[w_{V^\perp}]]| \right) = \sum_i \lambda_i(z) \left( |\pi_V[\pi_{V_i^\perp}[w_V]]| + |\pi_V[\pi_{V_i}[w_{V^\perp}]]| \right). \quad (4.67)$$

Since  $G$  is the Lipschitz graph of  $g$  over  $V$  with  $\|\nabla g\| \leq c\delta'$ , then  $\|\pi_{V^\perp}[w]\| \leq c\delta'$ . Then, by Lemma 4.4, we have

$$|\pi_V(A) - \pi_V[w]| \leq c \sum_i \lambda_i(z) (\delta^2 + \delta\delta'). \quad (4.68)$$

Summing up, since  $|\pi_V[w]| \leq |w| = 1$ , we obtain that

$$\left| |\nabla \sigma|_z[w]|^2 - 1 \right| = \left| |\pi_{V^\perp} \nabla \sigma|_z[w]|^2 + |(\pi_V \nabla \sigma|_z[w] - \pi_V[w]) + \pi_V[w]|^2 - 1 \right| \leq \quad (4.69)$$

$$\leq c(\delta + \delta')^2 + |\pi_V[w]|^2 - 1 = c(\delta + \delta')^2 + |\pi_{V^\perp}[w]|^2 \leq c(\delta + \delta')^2. \quad (4.70)$$

□

**4.5. Pointwise Estimates on  $D$ .** We wish to see in this subsection how (3.4) implies pointwise estimates on  $D$ , which will be convenient in the proof of the generalized Reifenberg results. Indeed, the following is an almost immediate consequence of Remark 3.5:

**Lemma 4.13.** *Assume  $B_r(x) \subseteq B_2(0)$  is such that  $\mu(B_r(x)) \geq 4^k \epsilon_n r^k$  and  $\int_{B_{2r}(x)} D_\mu^k(y, 2r) d\mu(y) < \delta^2 (2r)^k$ . Then there exists  $c(n)$  such that  $D(x, r) < c\delta^2$ . In particular, if (3.4) holds then for every  $B_r(x) \subseteq B_2(0)$  such that  $\mu(B_r(x)) \geq 4^k \epsilon_n r^k$  we have that  $D(x, r) < c\delta^2$ .*

## 5. PROOF OF THEOREM 3.3: THE RECTIFIABLE-REIFENBERG

Here we carry out the proof of Theorem 3.3.

In the proof, we will fix the constant  $M = C_1(n) \leq 40^n$  and therefore a positive scale  $\rho(n, C_1(n)) = \rho(n) < 1$  according to Lemma 4.7. For convenience, we will assume that  $\rho = 2^{-q}$ ,  $q \in \mathbb{N}$ , so that we will be able to use the sum bounds (3.9) more easily.

The constant  $C_1(n)$  will be defined by the end of the proof, however it is enough to know that it can be taken to be  $C_1(n) = 40^n$ . Thus, the value of the parameter  $\rho$  will depend only on the dimension  $n$ , as already pointed out in (4.13). Let us also define the scales  $r_j = \rho^j$ .

**5.1. Weak Upper bound  $\lambda^k(S \cap B_r(x)) \leq C(n)r^k$ .** We start by proving a uniform upper bound weaker than (3.8), in particular we want to show that for all  $x \in \mathbb{R}^n$  and  $r > 0$ :

$$\lambda^k(S \cap B_r(x)) \leq C(n)r^k. \quad (5.1)$$

Once we have obtained this estimate, the stronger upper bound and the rectifiability will be almost corollaries of this proof.

Given the scale invariance of the quantities involved, we do not lose generality if we prove (5.1) only for  $x = 0$  and  $r = 1$ . The strategy for the proof is the following: first we prove that  $S$  has  $\sigma$ -finite  $k$ -dimensional Hausdorff measure, and use this information to build a suitable covering by balls with controlled  $\lambda^k$ -measure. Then we fix any  $A \in \mathbb{N}$  and  $r_A = \rho^A$  and show by induction on  $j = A, \dots, 0$  that the measure of quasi-balls  $\tilde{B}_{r_j}(y)$ , defined in (5.11), is bounded above as in (5.1). The definition of a quasi-ball will be such that for  $r = 1$  the quasi-ball  $\tilde{B}_1(x)$  agrees with the set  $S \cap B_1(x)$ , up to a set of measure zero, which will prove (5.1). In order to prove the measure statement on the quasi-balls, we will use a second downward induction on  $i = j, \dots, A$ , which is the technical heart of the construction.

**$\sigma$ -finiteness of the measure** As a first step towards the proof, we remark that  $S$  must have  $\sigma$ -finite  $k$ -dimensional Hausdorff measure. We will use this to reduce the proof to the case when  $\lambda^k(S) < \infty$ . Indeed, in order for (3.7) to be true, we need in particular

$$\int_{S \cap B_1(0)} D_S^k(x, 1) d\lambda^k(x) ds < c\delta^2. \quad (5.2)$$

Define for  $a > 0$  the sets  $S_a = \{x \in S \text{ s.t. } D_S^k(x, 1) \geq a\}$  and  $S_0 = \{x \in S \text{ s.t. } D_S^k(x, 1) = 0\}$ . Then evidently for all  $a > 0$ ,  $\lambda^k(S_a) < c\delta^2/a < \infty$ . Moreover, if  $x \in S_0$ , then either  $\lambda^k(S \cap B_1(x)) < \epsilon_n \leq 1$ , or up to

sets of  $k$ -measure zero,  $S \cap B_1(x)$  is contained in a  $k$ -dimensional plane, and therefore  $\lambda^k(S_0) < \infty$ . Since  $S = S_0 \cup_{i=1}^{\infty} S_{i-1}$ ,  $S$  has  $\sigma$ -finite  $k$ -dimensional measure, as claimed.

Now in order to prove the uniform bounds on  $\lambda^k(S)$ , we can therefore assume without loss of generality that  $\lambda^k(S) < \infty$ . Indeed, by the monotonicity of Remark 3.4, all subsets of  $S$  satisfy (3.7). Therefore if we show that (5.1) holds for all subsets of finite  $k$ -dimensional measure, then the estimate will hold also on  $S$ . Thus, we will assume throughout the remainder of the proof that  $\lambda^k(S) < \infty$ .

**Covering of the set  $S^*$**  Let  $S^* \subseteq S \cap B_1(0)$  be the set of points with controlled upper density, i.e.

$$S^* = \left\{ x \in S \cap B_1(0) \text{ s.t. } 2^{-k} \leq \theta^{*k}(S, x) \leq 1 \right\}. \quad (5.3)$$

By proposition 2.16,  $S \setminus S^*$  has zero  $k$ -dimensional measure, thus it is sufficient to give bounds on  $S^*$ .

We want to cover the set  $S^*$  by balls which have uniform upper and lower bounds on their Hausdorff measure and such that their best  $L^2$  plane is not too far away from their center. In order to achieve this, for all  $x \in S^*$  let  $r_x$  be such that

$$r_x = \rho^{n_x}, \quad \text{where } n_x \in \mathbb{N}, \quad n_x \geq 2, \quad (5.4)$$

$$\begin{aligned} \lambda^k(B_{\rho r_x}(x) \cap S) &\geq \frac{1}{2^{k+1}} \omega_k \rho^k (\rho r_x)^k, \\ \lambda^k(B_r(x) \cap S) &\leq 2\omega_k r^k, \quad \forall r \leq r_x \end{aligned} \quad (5.5)$$

Let  $p_x$  be the center of mass of  $B_{\rho r_x}(x) \cap S$  with respect to  $\lambda^k|_{B_{\rho r_x}(x) \cap S}$ . In particular, we have that  $p_x \in B_{\rho r_x}(x)$ . By Jensen's inequality, we have

$$\lambda^k(B_{\rho r_x}(x) \cap S) d(p_x, V(x, r_x))^2 \leq \int_{B_{\rho r_x}(x) \cap S} d(y, V(x, r_x))^2 \leq r_x^{k+2} D(x, r_x), \quad (5.6)$$

where the last inequality comes from Definition 3.1, (4.13) and  $\epsilon_n = (1000n)^{-7n^2}$ . Using Lemma 4.13 and (3.7) we therefore get that

$$\lambda^k(B_{\rho r_x}(x) \cap S) d(p_x, V(x, r_x))^2 \leq c r_x^{k+2} \delta^2. \quad (5.7)$$

In particular, using our lower bound on  $\lambda^k(B_{r_x/8}(x) \cap S)$  we obtain that  $d(p_x, V(x, r_x)) \leq c\delta r_x$ , which implies for  $\delta$  small enough that

$$d(x, V(x, r_x)) \leq (\rho + c\delta)r_x \leq r_x/100. \quad (5.8)$$

Now consider the open covering of  $S^*$  given by  $\cup_{x \in S^*} B_{r_x/5}(x)$ , and extract a countable Vitali subcovering. Thus

$$S^* \subseteq \bigcup_{x \in \tilde{S}} B_{r_x}(x) \cap S^*, \quad (5.9)$$

where if  $x, y \in \tilde{S}$ ,  $x \neq y$ , then  $B_{r_x/5}(x) \cap B_{r_y/5}(y) = \emptyset$ .

Fix any  $A \in \mathbb{N}$  and  $\bar{r} = \rho^A$ , and define the sets  $\tilde{S}_{\bar{r}} = \{x \in \tilde{S} \text{ s.t. } \bar{r} \leq r_x \leq r\}$ ,  $\tilde{S}_{\bar{r}} = \tilde{S}_{\bar{r}}^{\rho^2}$  and

$$S_{\bar{r}} = \bigcup_{x \in \tilde{S}_{\bar{r}}} B_{r_x}(x) \cap S^*. \quad (5.10)$$

It is clear that  $S_{\bar{r}} \nearrow S^*$  as  $\bar{r} \rightarrow 0$ , so if we have bounds on  $\lambda^k(S_{\bar{r}})$  which are independent of  $\bar{r}$ , we are done. Thus from here on we will consider  $\bar{r} = \rho^A$  to be positive and fixed.

**5.2. First induction: upwards.** We are going to prove inductively on  $j = A, \dots, 0$  that for all  $x \in \mathbb{R}^n$  and  $r_j = \rho^j \leq 1$  the measure of

$$\tilde{B}_r(x) = \bigcup_{y \in \tilde{S}_{\bar{r}}^{\rho r} \cap B_r(x)} B_{r_y}(y) \cap S^*, \quad (5.11)$$

is bounded by

$$\lambda^k(\tilde{B}_{r_j}(x)) \leq C_1(n)r_j^k \leq 40^n r_j^k. \quad (5.12)$$

Note that for  $j = A$  this bound follows from the definition of  $r_x$  and (5.5). We emphasize that by construction  $B_{r_y}(y) \cap S^*$  appears in the union in (5.11) only if  $\bar{r} \leq r_y \leq \rho r$ . However, given (5.4), we have the inclusion  $S_{\bar{r}} \cap B_1(0) \subseteq \tilde{B}_1(0)$ . Note also that  $\tilde{B}_r(x) \subseteq B_{r(1+\rho)}(x)$ .

The reason why we have to introduce and estimate  $\tilde{B}_{r_j}(x)$  instead of  $B_{r_j}(x)$  is that we have no a priori control of what happens inside any of the balls  $B_{r_y}(y)$ . The bounds (5.5) are valid only on each ball as a whole. However, since our primary goal is to estimate  $S_{\bar{r}} \cap B_1(0)$ , and we have the inclusion  $S_{\bar{r}} \cap B_1(0) \subseteq \tilde{B}_1(0)$ , there is no loss in this restriction.

**Rough estimate** Fix some  $j$ , and suppose that (5.12) holds for all  $\bar{r} \leq r_k \leq r_{j+1}$ .

Let us first observe that we can easily obtain a bad upper bound on  $\lambda^k(\tilde{B}_{\chi r_{j+1}}(x))$  for some fixed  $\chi > 1$ . Consider the points  $y \in \tilde{S}_{\bar{r}} \cap B_{\chi r_{j+1}}(x)$ , and divide them into two groups: the ones with  $r_y \leq r_{j+2}$  and the ones with  $r_y \geq r_{j+1}$ .

For the first group, cover them by balls  $B_{r_{j+1}}(z_i)$  such that  $B_{r_{j+1}/2}(z_i)$  are disjoint. Since there can be at most  $c(n, \chi)$  balls of this form, and for all of these balls the upper bound (5.12) holds, we have an induced upper bound on the measure of this set.

As for the points with  $r_y > r_{j+2}$ , by construction there can be only  $c(n, \chi)$  many of them, and we also have the bound  $r_y \leq \rho \chi r_{j+1}$ , which by (5.5) implies  $\lambda^k(\tilde{B}_{r_y}(y)) \leq c(n, \chi)r_{j+1}^k$ . Summing up the two contributions, we get the very rough estimate

$$\lambda^k(\tilde{B}_{\chi r_{j+1}}(x)) \leq C_2(n, \chi)r_{j+1}^k, \quad (5.13)$$

where  $C_2 \gg C_1$ . Note that, as long as the inductive hypothesis holds,  $C_1$  and  $C_2$  are independent of  $j$ . However, it is clear that successive repetitions of the above estimate will not lead to (5.12). Our goal therefore is to push down this estimate to arrive at the better constant of  $C_1$ , however it will be technically very convenient when applying the tools of Section 4 that we may assume the worse bound in the process.

**5.3. Second induction: downwards: outline.** Suppose that (5.12) is true for all  $x \in B_1(0)$  and  $i = j + 1, \dots, A$ . Fix  $x \in \mathbb{R}^n$ , and consider the set

$$\tilde{B} = \tilde{B}_{r_j}(x) \subseteq B_{(1+\rho)r_j}(x).$$

We are going to build by induction on  $i \geq j$  a sequence of smooth maps  $\sigma_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and smooth  $k$ -dimensional manifolds  $T_{j,i} = T_i$  which satisfy nine properties, which we will use to eventually conclude the

proof of (5.12) for  $\tilde{B}$ . Let us outline the inductive procedure now, and introduce all the relevant terminology. Everything described in the remainder of this subsection will be discussed more precisely over the coming pages. To begin with, we will have at the first step that

$$\begin{aligned}\sigma_j &= id, \\ T_{j,j} &= T_j = \tilde{V}(x, r_j) \cap B_{2r_j}(x) \subseteq \mathbb{R}^n,\end{aligned}\tag{5.14}$$

where  $\tilde{V}(x, r_j)$  is one of the  $k$ -dimensional affine subspaces which minimizes  $\int_{\tilde{B}_{r_j}(x)} d^2(y, \tilde{V}) d\lambda^k$ . Thus, the first manifold  $T_j$  is a  $k$ -dimensional affine subspace which best approximates  $\tilde{B}_{r_j}(x)$ . At future steps we can recover  $T_i$  from  $T_{i-1}$  and  $\sigma_i$  from the simple relation

$$T_i = \sigma_i(T_{i-1}).\tag{5.15}$$

We will see that  $\sigma_i$  is a diffeomorphism when restricted to  $T_{i-1}$ , and thus each additional submanifold  $T_i$  is also diffeomorphic to  $\mathbb{R}^k$ . As part of our inductive construction we will build at each stage a covering of  $T_i$  given by

$$B_{r_j}(T_i) \cap \tilde{B}_{r_j}(x) \sim \bigcup_{s=j}^i \left( \bigcup_{y \in I_b^s} \tilde{B}_{r_s}(y) \bigcup_{y \in I_f^s} (B_{r_s}(y) \cap S_{\bar{r}}) \right) \bigcup_{y \in I_g^i} \tilde{B}_{r_i}(y),\tag{5.16}$$

where given any two distinct balls  $B_1$  and  $B_2$  in the covering,  $B_1/5$  and  $B_2/5$  are disjoint. Here  $I_g$ ,  $I_b$ , and  $I_f$  represent the *good*, *bad*, and *final* balls in the covering. A final ball  $\tilde{B}_{r_i}(y)$  with  $y \in I_f^i$  is a ball such that  $y \in \tilde{S}$  and  $r_y = r_i$ , and the other balls in the covering are characterized as good or bad according to how much measure they carry. Good balls are those with large measure, bad balls the ones with small measure. More precisely, we have

$$\begin{aligned}\lambda^k(\tilde{B}_{r_i}(y)) &\geq \gamma_k r_i^k, & \text{if } y \in I_g^i, \\ \lambda^k(\tilde{B}_{r_i}(y)) &< \gamma_k r_i^k, & \text{if } y \in I_b^i,\end{aligned}\tag{5.17}$$

where  $\gamma_k = \omega_k 40^{-k}$ . We will see that, over each ball  $\tilde{B}_{r_s}(y)$  in this covering,  $T_i$  can be written as a graph over the best approximating subspace  $\tilde{V}(y, r_s)$  with good estimates.

Our goal in these constructions is the proof of (5.12) for the ball  $\tilde{B} = \tilde{B}_{r_j}(x)$ , and thus we will need to relate the submanifolds  $T_i$ , and more importantly the covering (5.16), to the set  $\tilde{B}$ . Indeed, this covering of  $T_i$  almost covers the set  $\tilde{B}$ , at least up to an excess set  $\tilde{E}_{i-1}$ . That is,

$$\tilde{B} \subseteq \tilde{E}_{i-1} \cup \bigcup_{s=j}^i \left( \bigcup_{y \in I_b^s} \tilde{B}_{r_s}(y) \bigcup_{y \in I_f^s} (B_{r_s}(y) \cap S_{\bar{r}}) \right) \bigcup_{y \in I_g^i} \tilde{B}_{r_i}(y).\tag{5.18}$$

We will see that the set  $\tilde{E}_{i-1}$  consists of those points of  $\tilde{B}$  which do not satisfy a uniform Reifenberg condition. Thus in order to prove (5.12) we will need to estimate the covering (5.16), as well as the excess set  $\tilde{E}_{i-1}$ .

Let us now outline the main properties used in the inductive construction of the mapping  $\sigma_{j+1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , and hence  $T_{j+1} = \sigma_{j+1}(T_j)$ . As is suggested in (5.16), it is the good balls and not the bad and final balls which are subdivided at further steps of the induction procedure. In order to better understand this construction

let us begin by analyzing the good balls  $\tilde{B}_{r_i}(y)$  more carefully. On each such ball we may consider the best approximating  $k$ -dimensional subspace  $\tilde{V}(y, r_i)$ . Since  $\tilde{B}_{r_i}(y)$  is a good ball, one can check that most of  $\tilde{B} \cap \tilde{B}_{r_i}(y)$  must satisfy a uniform Reifenberg and reside in a small neighborhood of  $\tilde{V}(y, r_i)$ . We denote those points which don't by  $\tilde{E}(y, r_i)$ , see (5.33) for the precise definition. Then we can define the next step of the excess set by

$$\tilde{E}_i = \tilde{E}_{i-1} \bigcup_{y \in I_g^i} \tilde{E}(y, r_i). \quad (5.19)$$

Thus our excess set represents all those points which do not lie in an appropriately small neighborhood of the submanifolds  $T_i$ . With this in hand we can then find a submanifold  $T'_i \subseteq T_i$ , which is roughly defined by

$$T'_i \approx T_i \cap \bigcup_{y \in I_g^i} \tilde{B}_{r_i}(y), \quad (5.20)$$

see (5.47) for the precise inductive definition, such that

$$\tilde{B} \subseteq \tilde{E}_i \cup \bigcup_{s=j}^i \left( \bigcup_{y \in I_b^s} \tilde{B}_{r_s}(y) \bigcup_{y \in I_f^s} \tilde{B}_{r_s}(y) \right) \bigcup B_{r_{i+1}/4}(T'_i) \equiv \tilde{R}_i \bigcup B_{r_{i+1}/4}(T'_i), \quad (5.21)$$

where  $\tilde{R}_i$  represents our remainder term, and consists of those balls and sets which will not be further subdivided at the next stage of the induction. Now in order to finish the inductive step of the construction, we can cover  $B_{r_{i+1}/4}(T'_i)$  by some Vitali set

$$B_{r_{i+1}/4}(T'_i) \subseteq \bigcup_{y \in I} \tilde{B}_{r_{i+1}}(y), \quad (5.22)$$

where  $y \in I \subseteq T'_i$ . We may then decompose the ball centers

$$I^{i+1} = I_g^{i+1} \cup I_b^{i+1} \cup I_f^{i+1}, \quad (5.23)$$

based on (5.17). Now we will use Definition 4.10 and the best approximating subspaces  $V(y, r_{i+1})$  to build  $\sigma_{i+1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that

$$\text{supp}\{\sigma_{i+1} - Id\} \subseteq \bigcup_{y \in I_g^{i+1}} B_{3r_{i+1}}(y). \quad (5.24)$$

This completes the outline of the inductive construction.

**5.4. Second induction: downwards: detailed proof.** Let us now describe precisely the proof of this inductive construction which will lead to (5.12). For  $j \leq i \leq A$ , we will define a sequence of approximating manifolds  $T_i$  for the set  $S$  and a sequence of smooth maps  $\sigma_i$  such that

- (i)  $\sigma_j = id, T_{j,j} = T_j = \tilde{V}(x, r_j) \cap B_{2r_j}(x) \subseteq \mathbb{R}^n$ ,
- (ii)  $T_i = \sigma_i(T_{i-1})$ ,
- (iii) for  $y \in T_i$ ,

$$d(\sigma_{i+1}(y), y) \leq c\delta r_{i+1}, \quad (5.25)$$

and  $\sigma_{i+1}|_{T_i}$  is a diffeomorphism,

- (iv) for every  $y \in T_i$ ,  $T_i \cap B_{2r_i}(y)$  is the graph over some  $k$ -dimensional affine subspace of a smooth function  $f$  satisfying

$$\frac{\|f\|_\infty}{r_i} + \|\nabla f\|_\infty \leq c\delta, \quad (5.26)$$

As outlined before, the manifolds  $T_i$  will be good approximations of the set  $S$  up to some “excess” set of small measure. Moreover, we will also introduce the concept of *good*, *bad* and *final* balls (whose centers will be in the sets  $I_g^i$ ,  $I_b^i$  and  $I_f^i$ ), a *remainder set*  $R^i$ , and the manifolds  $T'_i \subseteq T_i$ . Before giving the precise definitions (which are in equations (5.44), (5.36), (5.34) and (5.47) respectively), let us group here all the properties that we will need (and prove) for these objects, so that the reader can always come back to this page to have a clear picture of what are the objectives of the proof.

- (v) for every  $i \geq j+1$  and  $y \in I_g^i$ ,  $d(y, \tilde{V}(y, r_i)) \leq c\delta r_i$ , the set  $T_i \cap B_{1.5r_i}(y)$  is the graph over  $\tilde{V}(y, r_i)$  of a smooth function  $f$  satisfying (5.26), where  $\tilde{V}(x, r)$  is one of the  $k$ -dimensional affine subspaces minimizing  $\int_{\tilde{B}_r(x)} d^2(y, \tilde{V}) d\lambda^k$ ,
- (vi) for all  $i$ , we have the inclusion

$$\tilde{B} \subseteq B_{r_i/2}(T'_{i-1}) \bigcup \tilde{R}^{i-1}, \quad (5.27)$$

- (vii) for every  $i \geq j+1$  and all  $y \in I_g^i$ , the set  $T'_{i-1} \cap B_{2r_i}(y)$  is a Lipschitz graph over the plane  $\tilde{V}(y, r_i)$  with  $d(y, \tilde{V}(y, r_i)) \leq c\delta r_i$ ,

The last two properties needed are the key for the final volume estimates:

- (viii) we can estimate

$$\lambda^k(\sigma_i^{-1}(T'_i)) + [\#(I_b^i) + \#(I_f^i)] \omega_k(r_i/10)^k \leq \lambda^k(T'_{i-1}), \quad (5.28)$$

- (ix) we can estimate the excess set by

$$\lambda^k(\tilde{E}(y, r_i)) r_{i+1}^2 \leq C(n) D_S^k(y, 2r_i). \quad (5.29)$$

**Inductive definitions** First of all, note that we can assume without loss of generality that

$$\lambda^k(\tilde{B}_{r_j}(x)) \geq \gamma_k r_j^k, \quad (5.30)$$

otherwise there's nothing to prove. With this hypothesis, we start our inductive construction by setting

$$I_g^j = \{x\}, \quad I_b^j = I_f^j = \emptyset, \quad T^j = \tilde{V}(x, r_j) \quad T'_j = T_j, \quad \sigma_j = id. \quad (5.31)$$

**Excess set.** Let us begin by describing the construction of the excess set. Fix any  $y$  and  $r_i \geq \rho^{-1}\bar{r} = \rho^{A-1}$ , and assume that  $\tilde{B}_{r_i}(y)$  satisfies  $\lambda^k(\tilde{B}_{r_i}(y)) \geq \gamma_k r_i^k$ .

Thus define  $\tilde{V}(y, r_i)$  to be (one of) the  $k$ -dimensional plane minimizing  $\int_{\tilde{B}_{r_i}(y)} d(y, \tilde{V})^2 d\lambda^k$ , and define also the excess set to be the set of points which are some definite amount away from the best plane  $\tilde{V}$ . Precisely,

$$E(y, r_i) = B_{r_i}(y) \setminus B_{r_{i+1}/4}(\tilde{V}), \quad \tilde{E}(y, r_i) = \bigcup_{y \in \tilde{S}_{\bar{r}}^{r_i+2} \cap E(y, r_i)} B_{r_y}(y) \cap S^*. \quad (5.32)$$



The points in  $\tilde{E}$  are in some sense what prevents the set  $S$  from satisfying a uniform Reifenberg condition on  $B_{r_i}(y)$ . By construction, all points in  $\tilde{E}$  have a uniform lower bound on the distance from  $\tilde{V}$ , so that if we assume  $\lambda^k(\tilde{B}_{r_i}(y)) \geq \gamma_k r_i^k$ , i.e.  $B_{r_i}(y)$  is a good ball, then we can estimate

$$\begin{aligned} \int_{\tilde{B}_{r_i}(y) \setminus \tilde{E}(y, r_i)} d(y, \tilde{V}(y, r_i))^2 d\lambda^k(y) + \lambda^k(\tilde{E}(y, r_i))(r_{i+1}/5)^2 &\leq \int_{\tilde{B}_{r_i}(y)} d(y, \tilde{V}(y, r_i))^2 d\lambda^k(y) \\ &\leq \int_{\tilde{B}_{r_i}(y)} d(y, V(y, r_i))^2 d\lambda^k(y) \leq \int_{B_{(1+\rho)r_i}(y)} d(y, V(y, (1+\rho)r_i))^2 \\ &= ((1+\rho)r_i)^{k+2} D_S^k(y, (1+\rho)r_i) \leq C(n)r_i^{k+2} D_S^k(y, 2r_i). \end{aligned} \quad (5.33)$$

**Good, bad and final balls, and remainder set** Inductively, let us define the remainder set to be the union of all the previous bad balls, final balls, and the excess sets:

$$R^i = \bigcup_{s=j}^i \left( \bigcup_{y \in I_b^s} B_{r_s}(y) \bigcup_{y \in I_f^s} B_{r_s}(y) \bigcup_{y \in I_g^s} E(y, r_s) \right), \quad \tilde{R}^i = \bigcup_{s=j}^i \left( \bigcup_{y \in I_b^s} \tilde{B}_{r_s}(y) \bigcup_{y \in I_f^s} (B_{r_s}(y) \cap S_{\bar{r}}) \bigcup_{y \in I_g^s} \tilde{E}(y, r_s) \right). \quad (5.34)$$

The set  $\tilde{R}^i$  represents everything we want to throw out at the inductive stage of the proof. We will see later in the proof how to estimate this remainder set itself.

Note that for  $i = j$ , i.e. at the first step of the induction, we have

$$R^j = E(x, r_j), \quad \tilde{R}^j = \tilde{E}(x, r_j). \quad (5.35)$$

Now consider the balls in the covering outside the remainder set, and separate the balls with radius  $\geq r_{i+1}$  from the others by defining for  $y \in I_g^i$  the sets

$$I_f^{i+1}(y) = \left\{ z \in \left( \tilde{S}_{\bar{r}} \setminus R^i \right) \cap B_{r_i}(y) \text{ s.t. } r_z = r_{i+1} \right\}, \quad (5.36)$$

and

$$J^{i+1}(y) = \left\{ z \in \left( \tilde{S}_{\bar{r}}^{r_{i+2}} \setminus R^i \right) \cap B_{r_i}(y) \cap B_{r_{i+1}/3}(\tilde{V}(y, r_i)) \right\}. \quad (5.37)$$

From this we can construct the sets

$$I_f^{i+1} = \bigcup_{y \in I_g^i} I_f^{i+1}(y) \quad \text{and} \quad J^{i+1} = \bigcup_{y \in I_g^i} J^{i+1}(y), \quad (5.38)$$

Note that by construction we have

$$S_{\bar{r}} \setminus \tilde{R}^i \subseteq \left( \bigcup_{z \in I_f^{i+1}} B_{r_{i+1}}(z) \bigcup_{z \in J^{i+1}} B_{r_z}(z) \right) \cap S_{\bar{r}}. \quad (5.39)$$

Let us now consider a minimal covering of (5.39) given by

$$S_{\bar{r}} \setminus \tilde{R}^i \subseteq \bigcup_{z \in I_f^{i+1}} B_{r_{i+1}}(z) \bigcup_{z \in I} B_{r_{i+1}}(z), \quad (5.40)$$

where  $I \subseteq T'_i$ , and for any  $p \neq q \in I_f^{i+1} \cup I$ ,  $B_{r_{i+1}/5}(p) \cap B_{r_{i+1}/5}(q) = \emptyset$ . Note that this second property is true by definition for  $p, q \in I_f^{i+1}$ , we only need to complete this partial Vitali covering with other balls of the same size. To be precise, note that by (5.34) and (5.32)

$$S_{\bar{r}} \setminus R^i \setminus \bigcup_{z \in I_f^{i+1}} B_{4r_{i+1}/5}(z) \subset \left( \bigcup_{y \in I_g^i} B_{r_{i+1}/4}(\tilde{V}(y, r_i)) \right) \cap \left( \bigcup_{y \in I_g^i} B_{r_i}(y) \right) \cap \left( \bigcup_{z \in I_f^{i+1}} B_{4r_{i+1}/5}(z) \right)^c. \quad (5.41)$$

Take a finite covering of this last set by balls  $\{B_{r_{i+1}/3}(y)\}_{y \in Y}$ . Note that we can pick

$$Y \cap \left( \bigcup_{z \in I_f^{i+1}} B_{r_{i+1}}(z) \right) = \emptyset. \quad (5.42)$$

Since  $T'_i$  is locally a Lipschitz graph over  $\tilde{V}(y, r_i)$  with (5.26), we can choose  $Y \subset T'_i$ . Moreover, since we have the inclusion  $S_{\bar{r}} \setminus \tilde{R}^i \subseteq \bigcup_{y \in I_g^i} \tilde{B}_{r_i}(y)$ , we can also choose  $Y \subset \bigcup_{y \in I_g^i} B_{r_i}(y)$ .

Consider a Vitali subcovering of this set, denote  $I$  the set of centers in this subcovering. Such a subcovering will have the property that the balls  $\{B_{r_{i+1}/3}(y)\}_{y \in I}$  will be pairwise disjoint. These balls will also be disjoint from  $\bigcup_{z \in I_f^{i+1}} B_{r_z/5}(z)$  by (5.42). The (finite version of) Vitali covering theorem ensures that  $\bigcup_{y \in I} B_{r_{i+1}}(y)$  will cover the whole set in (5.41).

Now by construction of  $I_f$  and the remainder set, all the balls  $\{B_{r_x}(x)\}_{x \in \bar{S}_{\bar{r}}}$  with  $r_x \geq r_{i+1}$  have already been accounted for. This means that

$$S_{\bar{r}} \setminus \tilde{R}^i \setminus \bigcup_{z \in I_f^{i+1}} B_{r_{i+1}}(z) \subset \bigcup_{y \in I} \tilde{B}_{r_{i+1}}(y), \quad (5.43)$$

as desired.

We split the balls with centers in  $I$  into two subsets, according to how much measure they carry. In particular, let

$$I_g^{i+1} = \left\{ y \in I \text{ s.t. } \lambda^k(\tilde{B}_{r_{i+1}}(y)) \geq \gamma_k r_{i+1}^k \right\}, \quad I_b^{i+1} = \left\{ y \in I \text{ s.t. } \lambda^k(\tilde{B}_{r_{i+1}}(y)) < \gamma_k r_{i+1}^k \right\}. \quad (5.44)$$

**5.5. Map and manifold structure.** Let  $\{\lambda_s^{i+1}\} = \{\lambda_s\}$  be a partition of unity such that for each  $y_s \in I_g^{i+1}$

- $\text{supp}(\lambda_s) \subseteq B_{3r_{i+1}}(y_s)$
- for all  $z \in \bigcup_{y_s \in I_g^{i+1}} B_{2r_{i+1}}(y_s)$ ,  $\sum_s \lambda_s(z) = 1$
- $\max_s \|\nabla \lambda_s\|_\infty \leq C(n)/r_{i+1}$ .

For every  $y_s \in I_g^{i+1}$ , let  $\tilde{V}(y_s, r_{i+1})$  to be (one of) the  $k$ -dimensional subspace that minimizes  $\int_{\tilde{B}_{r_{i+1}}(y_s)} d(z, V)^2 d\lambda^k$ .

By Remark 3.4 and by (5.11), we can estimate

$$r_{i+1}^{-k-2} \int_{\tilde{B}_{y_s}(r_{i+1})} d(z, \tilde{V}(y_s, r_{i+1}))^2 d\lambda^k(z) \leq (1 + \rho)^{k+2} D(y_s, (1 + \rho)r_{i+1}). \quad (5.45)$$

Let  $p_s \in B_{(1+\rho)r_{i+1}}(y_s)$  be the center of mass of  $\lambda^k|_{\tilde{B}_{r_{i+1}}(y_s)}$ . It is worth observing that  $p_s \in \tilde{V}_{y_s, r_{i+1}}$ .

Define the smooth function  $\sigma_{i+1} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  as in Definition 4.10, i.e.,

$$\sigma_{i+1}(x) = x + \sum_s \lambda_s^{i+1}(x) \pi_{\tilde{V}(y_s, r_{i+1})^\perp} (p_s - x). \quad (5.46)$$

With this function, we can define the sets

$$T_{i+1} = \sigma_{i+1}(T_i), \quad T'_{i+1} = \sigma_{i+1} \left( T_i \setminus \bigcup_{y \in I_f^{i+1}} B_{r_y/6}(y) \bigcup_{y \in I_b^{i+1}} B_{r_{i+1}/6}(y) \right). \quad (5.47)$$

Fix any  $y \in I_g^{i+1}$ , and let  $z \in I_g^i$  be such that  $y \in B_{r_i}(z)$ . By induction,  $T_i \cap B_{10r_{i+1}}(y) \subseteq T_i \cap B_{1.5r_i}(z)$  is the graph of a  $C^1$  function over  $\tilde{V}(z, r_i)$ . Consider the points  $\{y_s\} = I_g^{i+1} \cap B_{5r_{i+1}}(y)$ . By construction and using an estimate similar to (5.8), it is easy to see that  $d(y_s, \tilde{V}(z, r_i)) \leq c\delta r_{i+1}$ , and so we can apply the estimates in Lemma 4.8 with  $M = C_1$  by the first induction. Using condition (3.7), we obtain that for all  $y_s$ :

$$r_{i+1}^{-1} d_H \left( \tilde{V}(z, r_i) \cap B_{r_{i+1}}(y_s), \tilde{V}(y_s, r_{i+1}) \cap B_{r_{i+1}}(y_s) \right) \leq c \left( D_S^k(y_s, (1+\rho)r_{i+1}) + D_S^k(z, (1+\rho)r_i) \right)^{1/2} \leq c(n, \rho, C_1) \delta. \quad (5.48)$$

This implies that, if  $\delta(n, \rho, C_1)$  is small enough,  $T_i \cap B_{10r_{i+1}}(y)$  is a graph also over  $\tilde{V}(y, r_{i+1})$  satisfying the same estimates as in (5.26), up to a worse constant  $c$ . That is, if  $\delta$  is sufficiently small, we can apply Lemma 4.12 and prove induction point (v).

It is important to notice that on  $B_{1.5r_{i+1}}(y)$ , the bound on the Lipschitz constant of the graph is *independent* of the previous bound in the induction step by point (iii) in Lemma 4.12.

**Points (iii) and (iv)** Points (iii) and (iv) are proved with similar methods. We briefly sketch the proofs of these two points.

Let  $y \in T_{i+1}$ , and recall the function  $\psi_{i+1} \equiv 1 - \sum \lambda_s$ . If  $\psi_{i+1}|_{B_{2r_{i+1}}(y)}$  is identically 1, then  $\sigma_{i+1}|_{B_{2r_{i+1}}(y)} = id$ , and there is nothing to prove.

Otherwise, there must exist some  $z' \in I_g^{i+1} \cap B_{5r_{i+1}}(y)$ , and thus there exists a  $z \in I_g^i$  such that  $B_{3r_{i+1}}(y) \subseteq B_{1.5r_i}(z)$ . By point (iv) in the induction,  $T_i \cap B_{1.5r_i}(z)$  is a Lipschitz graph over  $\tilde{V}(z, r_i)$ . Proceeding as before, by the estimates in Lemma 4.8 and Lemma 4.12, we obtain that  $T_{i+1} \cap B_{2r_{i+1}}(y)$  is also a Lipschitz graph over  $\tilde{V}(z, r_i)$  with small Lipschitz constant, and that  $|\sigma_{i+1}(p) - p| \leq c\delta r_{i+1}$  for all  $p \in T_i$ .

Moreover,  $\sigma_{i+1}|_{T_i}$  is locally a diffeomorphism at scale  $r_{i+1}$ . From this we see that  $\sigma_{i+1}$  is a diffeomorphism on the whole  $T_i$ .

It is worth to remark a subtle point. In order to prove point (iii), we cannot use inductively (iii), we need to use point (iv). Indeed, as we have seen, given any  $z \in I_g^i$ , then  $T_i \cap B_{1.5r_i}(z)$  is a Lipschitz graph of a function  $f$  where  $|\nabla f| \leq c\delta$ , and this  $c$  is *independent* of the induction step we are considering by (iii) in Lemma 4.12. If we tried to iterate directly the bound given by (iii), the constant  $c$  would depend on the induction step  $i$ , and thus we could not conclude the estimate we want.

Now we turn our attention to the items (vi), (vii), (viii).

**Properties of the manifolds  $T'_i$**  Here we want to prove the measure estimate in (5.28). The basic idea is that

bad and final balls correspond to holes in the manifold  $T_i$ , and each of these holes carries a  $k$ -dimensional measure which is proportionate to the measure inside the balls. In particular, let  $B_{r_{i+1}}(y)$  be a bad or a final ball. In either case, we will see that  $y$  must be  $\sim r_{i+1}$ -close to  $T_i$ , which is a Lipschitz graph at scale  $r_i$ . This implies that  $\lambda^k(B_{r_i}(y) \cap T_i) \sim r_i^k$ , and thus we can bound the measure of a bad or final ball with the measure of the hole we have created on  $T_i$ .

In detail, point (vi) is an immediate consequence of the definition of  $\tilde{R}_i$ . As for point (vii), if  $y \in I_g^i$ , then by definition, there exists  $z \in I_g^{i-1}$  such that  $y \in B_{r_{i-1}}(z)$ , and  $y \notin R^{i-1}$ . This implies that  $y$  is far away from the balls we discard while building  $T'_{i-1}$ , in particular

$$B_{3r_i}(y) \cap \left( \bigcup_{y \in I_f^{i-1}} B_{r_{y/6}}(y) \bigcup_{I_b^{i-1}} B_{r_{i-1}/6}(y) \right) = \emptyset \quad (5.49)$$

This proves that  $T_{i-1} \cap B_{2r_i}(y) = T'_{i-1} \cap B_{2r_i}(y)$ , and in turn point (vii).

In order to prove the volume measure estimate, consider that

$$T'_i \setminus \sigma_{i+1}^{-1}(T'_{i+1}) \subseteq \left( \bigcup_{y \in I_f^{i+1} \cup I_b^{i+1}} B_{r_{i+1}/6}(y) \right). \quad (5.50)$$

Note that the balls in the collection  $\{B_{r_{i+1}/5}(y)\}_{y \in I_f^{i+1} \cup I_b^{i+1}}$  are pairwise disjoint. Pick any  $y \in I_b^{i+1}$ , and let  $z \in I_g^i$  be such that  $y \in B_{r_i}(z)$ . By definition,  $y \in T'_i$  and  $\lambda^k(\tilde{B}_{r_{i+1}}(y)) < \gamma k r_{i+1}^k < 10^{-k} \omega_k r_{i+1}^k$ . Since  $y \notin R^i$ , by (5.34)  $B_{r_{i+1}/6}(y)$  is disjoint from the set

$$\bigcup_{s=j}^i \left( \bigcup_{y \in I_b^s} B_{r_s/5}(y) \bigcup_{y \in I_f^s} B_{r_s/5}(y) \right) \quad (5.51)$$

and thus  $T_i \cap B_{r_{i+1}/6}(y) = T'_i \cap B_{r_{i+1}/6}(y)$ . Moreover,  $T_i \cap B_{2r_i}(z)$  is a graph over  $\tilde{V}(z, r_i)$  with  $y \in T'_i \subset T_i$ , thus (for  $\delta \leq \delta_0(n)$  small enough)

$$\lambda^k(T'_i \cap B_{r_{i+1}/6}(y)) \geq \omega_k 7^{-k} r_{i+1}^k. \quad (5.52)$$

A similar estimate holds for the final balls. The only difference is that if  $y \in I_f^{i+1}$ , then it is not true in general that  $y \in T_i$ . However, by construction of the balls  $B_{r_y}(y)$ , using an argument similar to the one in the proof of (5.7), we obtain that  $d(y, \tilde{V}(z, r_i)) \leq c\delta r_i + r_{i+1}/10$ . Given (5.26), we can conclude

$$\lambda^k(T'_i \cap B_{r_{i+1}/7}(y)) \geq \omega_k 10^{-k} r_{i+1}^k. \quad (5.53)$$

Now it is evident from the definition of  $T'_i$  that

$$\lambda^k(\sigma_i^{-1}(T'_i)) + \left[ \#(I_b^i) + \#(I_f^i) \right] \omega_k (r_i/10)^k \leq \lambda^k(T'_{i-1}). \quad (5.54)$$

**5.6. Volume estimates on the manifold part.** Here we want to prove that for every measurable  $\Omega \subseteq T_i$

$$\lambda^k(\sigma_{i+1}(\Omega)) \leq \lambda^k(\Omega) + c(n, \rho, C_1) \int_{S \cap B_{(1+\rho)r_j}(x)} D(p, 4r_{i+1}) d\lambda^k(p). \quad (5.55)$$

The main applications will be with  $\Omega = T_i \cap B_{2r_j}(x)$  and  $\Omega = T'_i \cap B_{2r_j}(x)$ . In order to do that, we need to analyze in a quantitative way the bi-Lipschitz correspondence between  $T_i$  and  $T_{i+1}$  given by  $\sigma_{i+1}$ .

As we already know,  $\sigma_{i+1} = id$  on the complement of the set  $G = \cup_{y \in I_g^{i+1}} B_{5r_{i+1}}(y)$ , so we can concentrate only on this set.

Using the same techniques as before, and in particular by Lemmas 4.8 and 4.12, we prove that for each  $y \in I_g^{i+1}$ , the set  $T_i \cap B_{5r_{i+1}}(y)$  is a Lipschitz graph over  $\tilde{V}(y, r_{i+1})$  with Lipschitz constant bounded by

$$c(n, \rho, C_1) \left( D(y, (1 + \rho)r_{i+1}) + \sum_{z \in I_g^i \cap B_{5r_i}(y)} D(z, (1 + \rho)r_i) \right)^{1/2} \quad (5.56)$$

In a similar manner, we also have that  $T_{i+1} \cap B_{5r_{i+1}}(y)$  is a Lipschitz graph over  $\tilde{V}(y, r_{i+1})$  with Lipschitz constant bounded by

$$c(n, \rho, C_1) \left( \sum_{z \in I_g^{i+1} \cap B_{10r_{i+1}}(y)} D(z, (1 + \rho)r_{i+1}) + \sum_{z \in I_g^i \cap B_{5r_i}(y)} D(z, (1 + \rho)r_i) \right)^{1/2}. \quad (5.57)$$

By the bi-Lipschitz estimates in the squash Lemma 4.12, we obtain that  $\sigma_{i+1}$  restricted to  $T_i \cap B_{5r_{i+1}}(y)$  is a bi-Lipschitz equivalence with bi-Lipschitz constant bounded by

$$L(y, 5r_{i+1}) \leq 1 + c \left( \sum_{z \in I_g^{i+1} \cap B_{10r_{i+1}}(y)} D(z, (1 + \rho)r_{i+1}) + \sum_{z \in I_g^i \cap B_{5r_i}(y)} D(z, (1 + \rho)r_i) \right) \quad (5.58)$$

In order to estimate this upper bound, we use an adapted version of (3.2) and the definition of good balls to write for all  $z \in I_g^{i+1}$

$$D(z, (1 + \rho)r_{i+1}) \leq c \int_{\tilde{B}_{r_{i+1}}(z)} D(p, 4r_{i+1}) d\lambda^k(p) \leq c(n, \rho, C_1) r_{i+1}^{-k} \int_{\tilde{B}_{r_{i+1}}(z)} D(p, 4r_{i+1}) d\lambda^k(p), \quad (5.59)$$

and a similar statement holds for  $z \in I_g^i$ . Since by construction any point  $x \in \mathbb{R}^n$  can be covered by at most  $c(n)$  different good balls at different scales, we can bound

$$L(y, 5r_{i+1}) \leq 1 + \frac{c(n, \rho, C_1)}{r_{i+1}^k} \int_{\tilde{B}_{5r_i}(y)} [D(p, 4r_{i+1}) + D(p, 4r_i)] d\lambda^k(p) \quad (5.60)$$

We can also bound estimate

$$D(p, 4r_{i+1}) + D(p, 4r_i) \leq c(n, \rho) D(p, 4r_i). \quad (5.61)$$

Now let  $P_s$  be a measurable partition of  $\Omega \cap G$  such that for each  $s$ ,  $P_s \subseteq B_{5r_{i+1}}(y_s)$ . By summing up the volume contributions of  $P_s$ , and since evidently  $\lambda^k(P_s) \leq c r_{i+1}^k$ , we get

$$\begin{aligned} \lambda^k(\sigma_{i+1}(\Omega)) &= \sum_s \lambda^k(\sigma_{i+1}(\Omega \cap P_s)) \leq \sum_s \lambda^k(P_s) \left( 1 + \frac{c}{r_{i+1}^k} \int_{\tilde{B}_{5r_i}(y_s)} D(p, 4r_i) d\lambda^k(p) \right) \leq \\ &\leq \lambda^k(\Omega) + c \int_{\cup_{y_s \in I_g^{i+1}} \tilde{B}_{5r_i}(y_s)} D(p, 4r_i) d\lambda^k(p) \leq \\ &\leq \lambda^k(\Omega) + c(n, \rho, C_1) \int_{B_{(1+\rho)r_i}(x) \cap \Omega} D(p, 4r_i) d\lambda^k(p). \end{aligned} \quad (5.62)$$

**5.7. Estimates on the excess set.** In this paragraph, we estimate the total measure of the excess set, which is defined by

$$\tilde{E}_T = \bigcup_{i=j}^A \bigcup_{y \in I_g^i} \tilde{E}(y, r_i). \quad (5.63)$$

At each  $y$  and at each scale  $r_i$  such that  $\lambda^k(\tilde{B}_{r_i}(y)) \geq \gamma_k r_i^k$ , we have by (5.33) and (3.2)

$$\lambda^k(\tilde{E}(y, r_i)) \leq c(n, \rho) r_i^k D_S^k(y, 2r_i) \leq c(n, \rho) r_i^k \int_{B_{2r_i}(y) \cap \mathcal{S}} D_S^k(p, 4r_i) d\lambda^k(p) \leq c(n, \rho) \int_{B_{2r_i}(y) \cap \mathcal{S}} D_S^k(p, 4r_i) d\lambda^k(p) \quad (5.64)$$

Now by construction of the good balls, there exists a constant  $c(n)$  such that at each step  $i$ , each  $x \in \mathbb{R}^n$  belongs to at most  $c(n)$  many balls of the form  $\{B_{2r_i}(y)\}_{y \in I_g^i}$ . Thus for each  $i \geq j$ , we have

$$\sum_{y \in I_g^i} \lambda^k(\tilde{E}(y, r_i)) \leq c(n, \rho) \int_{\bigcup_{y \in I_g^i} B_{2r_i}(y)} D_S^k(p, 4r_i) d\lambda^k(p) \leq c(n, \rho) \int_{B_{2r_j}(x) \cap \mathcal{S}} D_S^k(p, 4r_i) d\lambda^k(p). \quad (5.65)$$

If we sum over all scales, we get

$$\lambda^k(\tilde{E}_T) \leq c(n, \rho) \sum_{i=j}^A \int_{B_{2r_j}(x) \cap \mathcal{S}} D_S^k(p, 4r_i) d\lambda^k(p). \quad (5.66)$$

Since  $\rho = 2^{-q}$ , it is clear that

$$\lambda^k(\tilde{E}_T) \leq c(n, \rho) \sum_{i=j}^A \int_{B_{2r_j}(x) \cap \mathcal{S}} D_S^k(p, 2^{2-qi}) d\lambda^k(p) \leq c(n, \rho) \delta r_j^k, \quad (5.67)$$

since the sum in the middle is clearly bounded by (3.9).

This estimate is exactly what we want from the excess set.

**5.8. Completion of the weak upper bounds.** By adding (5.55), with  $\Omega \equiv \sigma_{i+1}^{-1}(T'_{i+1})$ , and (5.28), we prove that for all  $i = j, \dots, A+1, \dots$

$$\lambda^k(T'_{i+1}) + [\#(I_b^i) + \#(I_f^i)] \omega_k(r_i/10)^k \leq \lambda^k(T'_i) + c(n, \rho, C_1) \int_{B_{4r_j}(y) \cap \mathcal{S}_{\bar{r}}} D(p, 2r_i) d\lambda^k(p). \quad (5.68)$$

Adding the contributions from all scales, by (3.9) we get

$$\begin{aligned} & \lambda^k(T'_{i+1}) + \sum_{s=j}^i [\#(I_b^s) + \#(I_f^s)] \omega_k(r_s/10)^k \leq \\ & \leq \lambda^k(T_j \cap B_{2r_j}(x)) + c(n, \rho, C_1) \sum_{s=j}^i \int_{B_{2r_j}(x) \cap \mathcal{S}_{\bar{r}}} D(p, 2r_s) d\lambda^k(p) \leq \\ & \leq \lambda^k(T_j \cap B_{2r_j}(x)) [1 + c(n, \rho, C_1) \delta^2], \end{aligned} \quad (5.69)$$

where in the last line we estimated  $\lambda^k(T_j \cap B_{2r_j}(x)) \sim r_j^k$ , since  $T_j$  is a  $k$ -dimensional subspace, and we bounded the sum using (3.9).

In the same way, we can also bound the measure of  $T_i$  by

$$\lambda^k(T_{i+1}) \leq \lambda^k(T_j \cap B_{2r_j}(x)) \left[1 + c(n, \rho, C_1)\delta^2\right] \leq c(n)r_j^k. \quad (5.70)$$

**Upper estimates for  $\lambda^k$ .** Given the definition of  $\tilde{S}_{\bar{r}}$ , all the balls  $B_{r_y}(y)$  inside this set have  $r_y \geq r_A$ . Thus by (5.27)

$$\tilde{B} = \tilde{B}_{r_j}(x) \subseteq \tilde{R}^A. \quad (5.71)$$

In particular, this and the estimates in (5.67) and (5.69) imply

$$\begin{aligned} \lambda^k(\tilde{B}) &\leq \sum_{s=j}^A \left[ \gamma_k \#(I_b^s) + 2\omega_k \#(I_f^s) \right] (r_s)^k + \lambda^k(\tilde{E}_T) \leq \\ &\leq C_3(n) \sum_{s=j}^i \left[ \#(I_b^s) + \#(I_f^s) \right] \omega_k (r_s/10)^k + \lambda^k(\tilde{E}_T) \leq C_3(n)(1 + c(n, \rho, C_1)\delta)r_j^k. \end{aligned} \quad (5.72)$$

In this last estimate, we can fix  $C_1(n) = 2C_3(n) \leq 40^n$ , and  $\rho(n, C_1)$  according to Lemma 4.7. Now, it is easy to see that if  $\delta(n, \rho, C_1)$  is sufficiently small, then

$$\lambda^k(\tilde{B}) \leq C_1(n)r_j^k, \quad (5.73)$$

which finishes the proof of the downward induction, and hence the actual ball estimate (5.1).

**5.9. Rectifiability,  $W^{1,p}$  and improved measure bounds.** We have now proved a mass bound for the set  $S$ . In this subsection, we wish to improve this mass bound, as well as prove the rectifiability conditions on  $S$ . At several stages we will have to repeat the arguments of the upward and downward inductions, which we will only sketch since the arguments will be almost verbatim, though in some cases technically much easier.

**Rectifiability and  $W^{1,p}$  estimates** Now that we have proved the bound (5.1), we can use a construction similar to the one just described to sharpen the upper estimate and prove rectifiability of  $\lambda^k|_S$ . The main difference with the previous case is that we do not need to be concerned any more with any fixed covering  $B_{r_y}(y)$  of our set, and so at every step we can estimate directly the measure  $\lambda^k|_S$  of a whole Euclidean ball  $B_r(z)$  without the need to limit our estimate to  $\tilde{B}_r(z)$ . For the same reason, we do not need to introduce the subspaces  $\tilde{V}(z, r)$  and the sets  $\tilde{E}, \tilde{R}$ , we will only deal with  $V(z, r), E, R$ . Moreover, since we don't have to stop our construction at any positive scale  $r_y$ , we do not need to introduce and study the set of final balls  $I_f^j$ . As a consequence, the construction we need here is technically less involved than the one used before.

However, as opposed to the previous construction, we are concerned about what happens at an infinitesimal scale, and in particular we want to have uniform estimates also for the limit  $\bar{r} \rightarrow 0$  of various quantities.

Fix any  $x \in \mathbb{R}^n$  and  $r > 0$  such that  $B_r(x) \subseteq B_1(0)$ . We are going to prove (3.8), i.e.

$$\lambda^k(S \cap B_r(x)) \leq (1 + \epsilon)\omega_k r^k, \quad (5.74)$$

where  $\epsilon > 0$  is the arbitrary constant chosen at the beginning of Theorem 3.3.

For convenience and wlog, we assume  $x = 0$  and  $r = 1$ . Using the same technique as before, we build a sequence of smooth maps  $\{\sigma_i\}_{i=0}^\infty$  on  $\mathbb{R}^n$  and manifolds  $T_i$  such that

- (i)  $\sigma_0 = id$ ,  $T_0 = V(x, r) = V(0, 1) \subseteq \mathbb{R}^n$  and  $\phi_i = \sigma_i \circ \sigma_{i-1} \circ \cdots \circ \sigma_0$ ,
- (ii)  $T_i = \sigma_i(T_{i-1}) = \phi_i(T_0)$  is a smooth  $k$ -dimensional submanifold of  $\mathbb{R}^n$ ,
- (iii) for  $y \in T_i$

$$d(\sigma_{i+1}(y), y) \leq c\delta r_{i+1}, \quad (5.75)$$

and  $\sigma_{i+1}|_{T_i}$  is a diffeomorphism. In a similar way for all  $y \in T_0$ ,  $d(\phi_i(y), y) \leq c\delta$  and  $\phi_i|_{T_0}$  is a diffeomorphism,

- (iv) for every  $y \in T_i$ ,  $T_i \cap B_{2r_i}(y)$  is the graph over some  $k$ -dimensional affine subspace of a smooth function  $f$  satisfying

$$\frac{\|f\|_\infty}{r_i} + \|\nabla f\|_\infty \leq c\delta, \quad (5.76)$$

In order to do so, we define inductively on  $i = 0, \dots, \infty$  a sequence of sets  $E(y, r_i)$ ,  $I_g^i$ ,  $I_b^i$ ,  $R^i$  and manifolds  $T'_i \subseteq T_i$  such that

- (v) for every  $i \geq 1$  and all  $y \in I_g^i$ , the set  $T'_{i-1} \cap B_{1.5r_i}(y)$  is a Lipschitz graph over the plane  $V(y, r_i)$  with  $d(y, V(y, r_i)) \leq c\delta r_i$
- (vi) for all  $i$ , we have the inclusion

$$S \cap B_1(0) \subseteq B_{r_i/2}(T'_{i-1}) \cup R^{i-1} \quad (5.77)$$

- (vii) for every  $i \geq j+1$  and all  $y \in I_g^i$ , the set  $T'_{i-1} \cap B_{2r_i}(y)$  is a Lipschitz graph over the plane  $V(y, r_i)$  with  $d(y, V(y, r_i)) \leq c\delta r_i$ ,
- (viii) we can estimate

$$\lambda^k(\sigma_i^{-1}(T'_i)) + \#(I_b^i) \omega_k(r_i/10)^k \leq \lambda^k(T'_{i-1}). \quad (5.78)$$

Up to minor modifications (actually simplifications), all these properties are proved in the same way as in the downward induction from the previous subsection.

The key extra-property we need is some form of control over the  $W^{1,p}$  norm of  $\phi_i$ . In particular we will prove inductively that for all  $j \in \mathbb{N}$

$$\|\nabla \phi_j\|_{L^p}^p = \int_{T_0 \cap B_2(0)} |\nabla \phi_j|^p d\lambda^k \leq 1 + c(n)2^p \delta^2, \quad (5.79)$$

where here  $\lambda^k$  is also Lebesgue measure on  $\mathbb{R}^k$ .

As we will see in (5.98), in order for this estimate to work, we will need to choose  $\delta(n, \rho, C_1, p)$  small enough, and in particular  $\delta \rightarrow 0$  as  $p \rightarrow \infty$ .

**Proof of the first points** As mentioned before, the proof of items (i)-(viii) can be carried out in the same way as in Section 5.4, and we take from this section also the definitions of the sets  $I_g^i$ ,  $I_b^i$ ,  $E(x, r)$  and  $R^i$ , up



to replacing  $\tilde{B}_r(y)$  with  $B_r(y)$  and  $\tilde{V}(y, r)$  with  $V(y, r)$ . Recall also that in this case we have no final balls, so we can just assume that  $I_f^i = \emptyset$  for all  $i$ , and that in this case we have  $x = 0$ ,  $r = 1$  and so  $j = 0$ .

For convenience, we recall the definition of the excess and remainder sets:

$$E(y, r_i) = B_{r_i}(y) \setminus B_{r_{i+1}/4}(V(y, r_i)), \quad R^i = \bigcup_{s=0}^i \left( \bigcup_{y \in I_b^s} B_{r_s}(y) \bigcup_{y \in I_g^s} E(y, r_s) \right). \quad (5.80)$$

We also introduce the notation

$$E^i = \bigcup_{s=0}^i \bigcup_{y \in I_g^s} E(y, r_s), \quad R^i = E^i \bigcup_{s=0}^i \bigcup_{y \in I_b^s} B_{r_s}(y). \quad (5.81)$$

We briefly sketch again the main steps in the construction. Assuming wlog that  $B_1(0)$  is a good ball, i.e., that  $\lambda^k(B_1(0)) \geq \gamma_k$ , we first estimate the excess set on this ball. Since this set is the set of points which are some definite amount away from the best plane  $V(0, 1)$ , the definition of  $D$  immediately gives the following estimate, similar to (5.33):

$$\lambda^k(E(0, 1))(\rho/5)^2 \leq D_S^k(0, 1). \quad (5.82)$$

Then we cover the non-excess part with a Vitali-type covering by balls  $B_\rho(x_i)_{i \in I}$  centered on the plane  $V(0, 1)$ . We classify the balls in this covering into good and bad balls, according to how much mass they carry.  $B_r(x)$  is a good ball if  $\lambda^k(S \cap B_r(x)) \geq \gamma_k r^k$ , otherwise it's a bad ball.

A good ball carries enough measure to apply Lemma 4.8, and compare the best subspace  $V(0, 1)$  with the new best subspace  $V(x_i, \rho)$ .

We set  $\sigma_1$  to be the map defined in 4.10, i.e., an interpolation among all the projections onto  $\{V(x_i, \rho)\}_{i \in I_g^1}$ , and we also set  $T_1 = \sigma_1(T_0) = \sigma_1(V(0, 1) \cap B_{1+c\delta}(0))$ . By the squash lemma 4.12,  $\sigma_1|_{T_0}$  is a smooth diffeomorphism, and  $T_1$  is locally at scale  $r_1 = \rho^1$  the Lipschitz graph of a function with small Lipschitz bounds. Moreover,  $\sigma_1$  is a bi-Lipschitz equivalence with quantitative estimates on the Lipschitz constant. In particular, we get that for all  $y \in B_1(0) \cap T_0$ , the following version of the estimates in (5.58) holds:

$$L(y, 5\rho) = \max \left\{ \|\nabla \sigma_1\|_{L^\infty(T_0 \cap B_{5\rho}(y))}, \|\nabla \sigma_1\|_{L^\infty(\sigma_1(T_0 \cap B_{5\rho}(y)))} \right\} \leq 1 + c \left( \sum_{z \in I_b^1 \cap B_{10\rho}(y)} D(z, \rho) + D(0, 1) \right) \quad (5.83)$$

In order to keep track of the measure inside the bad balls, we define the manifold “with holes”

$$T_1 = \sigma_1(T_0), \quad T'_1 = \sigma_1 \left( T_0 \setminus \bigcup_{z \in I_b^1} B_{\rho/6}(z) \right). \quad (5.84)$$

Since  $V(0, 1)$  is a  $k$ -plane, and by definition of bad balls, each “hole” in the manifold  $T'_1$  carries more measure than the corresponding bad ball which created it, giving us the estimate

$$\lambda^k(\sigma_1^{-1}(T'_1)) + \lambda^k \left( \bigcup_{z \in I_b^1} S \cap B_\rho(z) \right) \leq \lambda^k(\sigma_1^{-1}(T'_1)) + \#(I_b^1) \omega_k(\rho/9)^k \leq \lambda^k(T_0). \quad (5.85)$$

Now we repeat the construction inductively on the scales  $r_2 = \rho^2, r_3, \dots$ , and we obtain all the desired properties. Moreover, by summing (5.82) and its iterations, we obtain the following estimate for the total excess set:

$$\lambda^k(E^\infty) \leq c(n, \rho)\delta^2, \quad (5.86)$$

which is the equivalent of (5.67).

We conclude by noting the following:  $\sigma|_i(z) = id$  for all  $z \in T_0 \cap \left(\bigcup_{j=0}^i \bigcup_{z \in I_b^j} B_{r_i/2}(z)\right)$ . In other words, once a hole is created, it never changes. This implies that the iterations of (5.85) lead to

$$\lambda^k\left(T'_i \cap B_{1+c\delta}(0)\right) + 9^{-k} \gamma_k^{-1} \sum_{j=0}^i \sum_{y \in I_b^j} \lambda^k(S \cap B_{r_j}(y)) \leq \lambda^k(T_i \cap B_{1+c\delta}(0)). \quad (5.87)$$

**$W^{1,p}$  estimates** As part of the downward induction in the proof of (i)-(viii), let us define the maps  $\tau_{i,j} = \sigma_j \circ \sigma_{j-1} \circ \dots \circ \sigma_{i+1}$ , so that  $\tau_{i,j}(T_i) = T_{j+1}$  and  $\tau_{0,j} = \phi_j$ . We will prove inductively for  $i = j, \dots, 0$  that for all  $x \in T_i$ ,

$$r_i^{-k} \int_{B_{r_i}(x) \cap T_i} |\nabla \tau_{i,j}|^p d\lambda^k \leq 2C_1(n). \quad (5.88)$$

Here the integration is simply the integration on a smooth  $k$ -dimensional subset wrt the  $k$ -dimensional Lebesgue measure  $\lambda^k$ , and the gradient of the functions  $\sigma_i$  and  $\tau_{i,j}$  is the restriction of the gradient in  $\mathbb{R}^n$  to the corresponding manifold  $T_i$ .

Suppose that the statement is true for  $i+1$ . Consider a covering of this ball by  $\{B_{r_{i+1}}(y_s)\}$ , then by the chain rule

$$\int_{B_{r_i}(x) \cap T_i} |\nabla \tau_{i,j}|^p d\lambda^k(y) \leq \sum_s \int_{B_{r_{i+1}}(y_s) \cap T_i} |\nabla \tau_{i+1,j}|_{\sigma_i(y)}|^p |\nabla \sigma_i(y)|^p d\lambda^k(y) \leq C(n, p, \rho). \quad (5.89)$$

This gives us a first rough estimate.

In order to obtain a better estimate, we will prove by induction on  $s = i, \dots, j$  that

$$\int_{B_{r_i}(x) \cap T_i} |\nabla \tau_{i,j}|^p d\lambda^k(y) \leq \int_{B_s} |\nabla \tau_{s,j}|^p d\lambda^k(y) + c(n, \rho) 2^{2p} \sum_{t=i}^s \int_{B_{10r_t}(x) \cap S} D(p, 10r_t) d\lambda^k(p), \quad (5.90)$$

where  $B_s \equiv \tau_{i,s-1}(T_i \cap B_{r_i}(x)) \subseteq T_s \cap B_{(1+c\delta)r_i}(x)$ . Indeed, suppose that the statement is true for  $s$ , then we have

$$\int_{B_s} |\nabla \tau_{s,j}|^p d\lambda^k(y) = \int_{B_s} |\nabla \tau_{s+1,j}|_{\sigma_s(y)}|^p |\nabla \sigma_s(y)|^p d\lambda^k(y). \quad (5.91)$$

Since  $\sigma_i$  is a diffeomorphism, we can change the variables and write

$$\int_{B_s} |\nabla \tau_{s,j}|^p d\lambda^k(y) = \int_{B_{s+1}} |\nabla \tau_{s+1,j}|_y|^p |\nabla \sigma_s|_{\sigma_s^{-1}(y)}|^p \det J(\sigma_s) d\lambda^k(y). \quad (5.92)$$

Now consider the partial covering of this set given by  $\{B_{5r_s}(y_t)\}_{y_t \in I_b^s}$ . As we have seen before, outside of this set,  $\sigma_s$  is the identity, so we don't need to make any estimates on it.

Arguing in a manner verbatim to the proof of (5.55), we can prove bi-Lipschitz estimates for  $\sigma|_s$ . In particular, we can use the definition of good balls and the bounds in Lemma 4.8 to estimate in a quantitative

way the distance between best subspaces at nearby points and scales. This allows us to use the squash Lemma 4.12 and prove that  $\sigma_s$  restricted to  $T_s \cap B_{5r_{s+1}}(y_t)$  is a bi-Lipschitz map with

$$\left\| |\nabla \sigma_s|_{T_s} - 1 \right\|_{L^\infty(T_s \cap B_{5r_{s+1}}(y_t))} \leq \frac{c(n, \rho)}{r_s^k} \int_{B_{5r_s}(y_t) \cap \mathcal{S}} D(p, 10r_s) d\lambda^k(p). \quad (5.93)$$

Let  $\{P_t, Q\}$  be a measurable partition of  $B_s$  with  $P_t \subseteq B_{5r_{s+1}}(y_t)$  and  $Q \subseteq \cap B_{5r_{s+1}}(y_t)^c$ , then  $\{\sigma_s(P_t), \sigma_s(Q)\}$  is a measurable partition of  $B_{s+1}$  with  $\sigma_s(Q) = Q$ . So we get

$$\begin{aligned} \int_{B_s} |\nabla \tau_{s,j}|^p d\lambda^k(y) &= \int_Q |\nabla \tau_{s,j}|^p d\lambda^k(y) + \sum_t \int_{\sigma(P_t)} |\nabla \tau_{s+1,j}|^p |\nabla \sigma_s|_{\sigma_s^{-1}(y)}|^p \det J(\sigma_s) d\lambda^k(y) \leq \\ &\leq \int_Q |\nabla \tau_{s+1,j}|^p d\lambda^k(y) + \sum_t \int_{\sigma(P_t)} |\nabla \tau_{s+1,j}|^p d\lambda^k \left( 1 + \frac{c}{r_{s+1}^k} \int_{B_{5r_s}(y_t) \cap \mathcal{S}} D(p, 10r_s) d\lambda^k(p) \right). \end{aligned} \quad (5.94)$$

Note that these estimates are basically the same as the ones in (5.62).

By the first induction, we have the upper bound  $\int_{P_s} |\nabla \tau_{s+1,j}|^p \leq c(n)r_{s+1}^k$ , and thus

$$\int_{B_s} |\nabla \tau_{s,j}|^p d\lambda^k(y) \leq \int_{B_{s+1}} |\nabla \tau_{s+1,j}|^p d\lambda^k(y) + c(n, \rho) 2^p \sum_t \int_{B_{5r_{s+1}}(y_t) \cap \mathcal{S}} D(p, 10r_s) d\lambda^k(p). \quad (5.95)$$

Since all points in  $\mathbb{R}^n$  are covered at most  $c(n, \rho)$  times by  $\{B_{5r_{s+1}}(y_t)\}_t$ , we can simply estimate

$$\int_{B_s} |\nabla \tau_{s,j}|^p d\lambda^k(y) \leq \int_{B_{s+1}} |\nabla \tau_{s+1,j}|^p d\lambda^k(y) + c(n, \rho) 2^p \int_{B_{10r_t}(x) \cap \mathcal{S}} D(p, 10r_s) d\lambda^k(p). \quad (5.96)$$

This proves (5.90). In order to complete the first induction, observe that

$$\int_{B_{r_i}(x) \cap T_i} |\nabla \tau_{i,j}|^p d\lambda^k(y) \leq \int_{B_s} |\nabla \tau_{j,j}|^p d\lambda^k(y) + c(n, \rho) 2^p \sum_{t=i}^j \int_{B_{2r_t}(x) \cap \mathcal{S}} D(p, 10r_t) d\lambda^k(p), \quad (5.97)$$

Since  $\tau_{j,j}$  is the identity on  $T_j$ , and given the measure bounds (5.70), we get the result if we choose  $\delta(n, \rho, p)$  such that the rhs is small enough, and in particular

$$c(n, \rho) 2^p \sum_{j=i}^A \int_{\mathcal{S} \cap B_{2r_j}(x)} D(p, 10r_{i-1}) d\lambda^k(p) \leq c(n, \rho) 2^p \delta^2 \leq C_1(n). \quad (5.98)$$

Note that, once the improved estimate (3.8) is proven (which will be done in the next subsection) it will likewise be possible to improve this  $W^{1,p}$  estimate to

$$\frac{1}{\omega_k r_i^k} \int_{B_{r_i}(x) \cap T_i} |\nabla \tau_{i,j}|^p d\lambda^k \leq 1 + \epsilon. \quad (5.99)$$

**Improved measure estimates** Consider the maps  $\phi_j : T_0 \rightarrow T_{j+1}$ . Since  $\phi_j$  has uniform  $W^{1,p}$  estimates and by point (iii), then  $\phi_j \rightarrow \phi_\infty$  uniformly, with

$$\int_{T_0 \cap B_2(0)} |\nabla \phi_\infty|^p \leq 2C_1(n). \quad (5.100)$$

By (5.77), we have the inclusion for all  $i$ :

$$S \subseteq E_i \bigcup_{j=0}^i B_{r_{i+1}/2}(T'_i) \bigcup_{y \in I_b^j} B_{r_j}(y), \implies S \cap B_1(0) \subseteq E_\infty \bigcup_{j=0}^\infty (T'_\infty \cap B_1(0)) \bigcup_{y \in I_b^j} B_{r_j}(y). \quad (5.101)$$

By (5.87) and (5.86), we obtain that

$$\lambda^k(S \cap B_1(0)) \leq \lambda^k(T_\infty) + \lambda^k(E_\infty) \leq \lambda^k(T_\infty) + c\delta^2. \quad (5.102)$$

In order to give better estimates on  $T_\infty$ , note that  $T_\infty \cap B_1(0) \subseteq \phi(T_0 \cap B_{1+c\delta}(0))$ . Define the function  $f : T_0 \rightarrow \mathbb{R}$  by

$$f(x) = \sum_{i=0}^\infty \sup_{B_{r_i}(\phi_i(x)) \cap T_i} \|\nabla \sigma_i - 1\|. \quad (5.103)$$

By the usual double-induction argument, see for example the proofs of (5.73) and (5.88), we can estimate

$$\int_{T_0 \cap B_{1+c\delta}(0)} f(x) d\lambda^k \leq c\delta^2. \quad (5.104)$$

Define the sets  $U_a \subseteq T_0$  by

$$U_a = \{x \in T_0 \cap B_{1+c\delta}(0) \text{ s.t. } f(x) > a\}. \quad (5.105)$$

By simple  $L^1$  estimates, we know that  $\lambda^k(U_a) \leq \frac{c\delta^2}{a}$ , and by Lemma 2.20 for  $p > k$  we get

$$\lambda^k(\phi_\infty(U_a)) \leq C(n, p) \|\nabla \phi_\infty\|_{L^p}^k \lambda^k(U_a)^{\frac{p-k}{p}} \leq C \left( \frac{c\delta^2}{a} \right)^{1-k/p}. \quad (5.106)$$

Note that, on the complement of  $U_a$ ,  $\phi_\infty$  is a Lipschitz function with  $\|\nabla \phi_\infty\|_\infty \leq e^a + c\delta$ . Indeed, take any  $x, y \in U_a^c$ , and let  $i$  be such that  $r_{i+1}/2 < |x - y| \leq r_i/2$ . We prove that

$$|\phi_\infty(x) - \phi_\infty(y)| \leq (e^a + c\delta) |x - y|. \quad (5.107)$$

By definition,  $\phi_\infty(z) = \tau_{i_\infty}(\phi_i(z))$ , and by (5.75), we easily get

$$|\tau_{i_\infty}(\phi_i(x)) - \phi_i(x)| + |\tau_{i_\infty}(\phi_i(y)) - \phi_i(y)| \leq c\delta r_i. \quad (5.108)$$

Moreover, the definition of  $f$  immediately implies a uniform Lipschitz condition on  $\phi_i$ . Indeed,

$$\|\nabla \phi_i\|_{L^\infty(B_{r_i}(x))} \leq \prod_{s=0}^i \|\nabla \sigma_s\|_{L^\infty(B_{r_s}(\phi_s(x)))} \leq \exp \left( \sum_{s=0}^i \log \left( \|\nabla \sigma_s\|_{L^\infty(B_{r_s}(\phi_s(x)))} \right) \right) \leq e^a. \quad (5.109)$$

Thus we get

$$|\psi(x) - \psi(y)| \leq c\delta |x - y| + e^a |x - y|, \quad (5.110)$$

as desired.

By choosing  $a(\epsilon)$  sufficiently small, and  $\delta(n, \rho, p, \epsilon)$  sufficiently small as well, we get

$$\lambda^k(T_\infty \cap B_1(0)) \leq \lambda^k(\phi_\infty(U_a) \cup \phi_\infty(U_a^c)) \leq \left(1 + \frac{1}{2}\epsilon\right) \omega_k. \quad (5.111)$$

Summing up all the estimates, we prove the sharpened upper bound

$$\lambda^k(S \cap B_1(0)) \leq \lambda^k(E_\infty) + \lambda^k(\phi_\infty(T_0 \cap B_{1+c\delta}(0))) \leq (1 + \epsilon) \omega_k. \quad (5.112)$$

**Rectifiability** As for the rectifiability, we can restrict ourselves to  $S^\star$  and consider a covering of this set made by balls  $\{B_{r_x}(x)\}_{x \in S^\star}$  such that  $\lambda^k(B_{r_x}(x) \cap S) \geq 2^{-k-1}\omega_k r_x^k$ .

If we prove that for all such balls there exists a subset  $G \subseteq S \cap B_{r_x}(x)$  which is rectifiable and for which  $\lambda^k(G) \geq \lambda^k(B_{r_x}(x) \cap S)/2$ , then an easy covering argument gives us the rectifiability of all of  $S$ .

By scale invariance, we assume for simplicity  $x = 0$  and  $r_x = 1$ . As we have seen before,

$$S \subseteq R^\infty \cup T_\infty. \quad (5.113)$$

Since  $T_\infty$  is the image of a  $W^{1,p}$  map with  $p > k$ , this set is rectifiable by Lemma 2.19. Moreover, by (5.87) and (5.111), we know that

$$\sum_{j=0}^i \sum_{y \in J_b^j} \lambda^k(S \cap B_{r_j}(y)) \leq 10^k \gamma_k \quad (5.114)$$

Given the estimates on the excess set given in (5.86), we obtain the lower bound

$$\lambda^k(S \cap T_\infty) \geq \lambda^k(S) - \lambda^k(R^\infty) \geq 2^{-k-1}\omega_k - \gamma_k 10^k - c\delta \geq 2^{-k-2}\omega_k, \quad (5.115)$$

which therefore completes the proof.

## 6. PROOF OF THEOREMS 3.2 AND 3.4

As it will be clear, the proofs of these theorems are simple modifications of the proof of Theorem 3.3, and actually from the technical point of view they are a lot simpler. For this reason, we will simply outline them, pointing out the main differences needed in these cases.

**6.1. Proof of Theorem 3.2: The  $W^{1,p}$ -Reifenberg.** The proof of this theorem is almost a corollary of Theorem 3.3 and the classic Reifenberg theorem 2.18. In addition to the upper volume bound proved in Theorem 3.3, we also need to prove a lower volume bound on each ball.

In fact, one could just trace through the argument of Theorem 3.3, and see inductively that there exists no bad balls to conclude this. Instead, we will use a standard argument (see for example [DT12, lemma 13.2]) to prove the lower bound on  $\lambda^k(S \cap B_r(x))$  given in (3.6) directly by using only the uniform Reifenberg condition.

**Lemma 6.1.** *Under the assumptions of Theorem 2.18, for all  $x \in S$  such that  $B_r(x) \subseteq B_1(0)$ ,*

$$\lambda^k(S \cap B_r(x)) \geq (1 - c\delta)\omega_k r^k. \quad (6.1)$$

*Proof.* By scale invariance, we assume  $x = 0$  and  $r = 1$ . The classic Reifenberg theorem proves that there exists a bi-Hölder continuous map  $\phi : L \rightarrow \mathbb{R}^n$  where  $L$  is a  $k$ -dimensional plane and

- (1)  $|\phi(x) - x| \leq c\delta$  for all  $x \in L$
- (2)  $\phi(x) = x$  for  $|x| \geq 1 + c\delta$
- (3)  $S \cap B_1(0) = \phi(L')$ , where  $B_{1-c\delta}(0) \cap L \subseteq L' \subseteq B_{1+c\delta}(0) \cap L$ .

Now let  $f = \pi_L \circ \phi : L \rightarrow L$ . This map is continuous and it is the identity outside  $B_{1+c\delta}(0)$ , and thus by topological reasons (degree theory) it is also surjective from  $L$  to itself.

In particular, the set  $A = B_{1-3c\delta}(0) \cap L$  is contained in the image of  $f$ . By the uniform Reifenberg condition,  $\pi_L^{-1}(A) \cap S \subseteq B_{1-2c\delta}(0)$ , and by the properties of  $\phi$ ,  $f^{-1}(A) = \phi^{-1}(\pi_L^{-1}(A) \cap S) \subseteq B_{1-c\delta}(0)$ . Thus  $\phi(f^{-1}(A)) \subseteq S$ .

Now, since  $\pi_L$  has Lipschitz constant 1, by a standard result (see [Fed69, 2.10.11])

$$\lambda^k(S) \geq \lambda^k(\phi(f^{-1}(A))) \geq \lambda^k(\pi_L \circ \phi(f^{-1}(A))) = \lambda^k(A) \geq (1 - 3c\delta)^k \omega_k r^k. \quad (6.2)$$

For  $\delta$  small enough, we have the thesis. □

The upper bound on  $\lambda^k(S)$  and the rectifiability are direct consequences of Theorem 3.3. The only thing left to prove is the statement about  $W^{1,p}$  bounds on the map. This can be obtained by simple modifications (actually simplifications) in the proof of Theorem 3.3.

As before, we denote by  $V(x, r)$  (one of) the  $k$ -dimensional subspace that minimizes  $\int_{B_r(x)} d(y, V)^2$ , and we set  $L(x, r)$  to be one of the  $k$ -dimensional subspaces satisfying the Reifenberg condition, i.e., a  $k$ -dimensional subspace such that  $d_H(S \cap B_r(x), L \cap B_r(x)) < \delta r$ .

First of all, note that for any  $x \in S$ ,  $r > 0$ ,  $\lambda^k(B_r(x)) > \omega_k r^k / 2$ , and so there are no bad balls in our covering. Thus, by Lemma 4.9, we can conclude that

$$d_H(V(x, r) \cap B_r(x), L(x, r) \cap B_r(x)) \leq c\delta. \quad (6.3)$$

As a consequence, for  $\delta(n, \rho)$  small enough, all the excess sets at all scales are empty. Indeed, by (5.32),

$$E(x, r) = B_r(x) \setminus B_{\rho r/3}(V(x, r)) \cap S \subseteq B_r(x) \setminus B_{(\rho/3-c\delta)r}(L(x, r)) \cap S = \emptyset. \quad (6.4)$$

Proceeding with the same construction as in Section 5.9, we obtain a sequence of maps  $\phi_j : T_0 \rightarrow \mathbb{R}^n$ ,  $\phi_j(T_0) \equiv T_{j+1}$  converging in  $W^{1,p}$  to some  $\phi_\infty$  such that for all  $i$ ,  $S \subseteq B_{r_i}(T_i) \subseteq B_{r_{i-1}}(T_{i-1})$ , thus proving that  $S \subseteq T_\infty \equiv \phi_\infty(T_0)$ .

**$W^{1,p}$  estimates for the inverse** Here we want to prove that the map  $\phi_\infty^{-1} : S \rightarrow T$  is also a  $W^{1,p}$  map with bounds. Note that  $S$  equipped with the Euclidean distance and the  $k$ -dimensional Hausdorff measure is a Ahlfors regular metric measure space, in the sense that there exists a  $C_1(n)$  such that for all  $x \in S \cap B_1(0)$  and  $r \leq 1$ ,

$$C_1^{-1} r^k \leq \lambda^k(S \cap B_r(x)) \leq C_1 r^k, \quad (6.5)$$

where  $B_r(x)$  is the usual Euclidean ball in  $\mathbb{R}^n$ . On such spaces there are several methods of defining the space of  $W^{1,p}$  maps, for instance as the closure of the Lipschitz functions under the  $W^{1,p}$  norm. For rectifiable spaces all such definitions are classically understood to be equivalent.

Note that since  $S$  is rectifiable, we can use the integration by substitution to write

$$\int_S f(z) d\lambda^k(z) = \int_{\phi_\infty^{-1}(S)} f(\phi_\infty(x)) J(\phi_\infty)|_x d\lambda^k(x), \quad (6.6)$$

which will allow us to easily study integrals on our rectifiable spaces.

The next Lemma tells us that the mapping  $\phi_\infty^{-1}$  is approximated by a sequence of Lipschitz maps whose gradients form a Cauchy sequence in  $L^p$  and have our desired estimates. In particular, the following Lemma will finish our proof of the  $W^{1,p}$ -Reifenberg result. For convenience, we introduce the notation  $\psi \equiv \phi_\infty^{-1}$ :

**Lemma 6.2.** *There exists a sequence of functions  $\psi_t$  such that*

- (1)  $\psi_t$  are Lipschitz functions in  $\mathbb{R}^n$ , with Lipschitz constant bounded by  $c(n)t$ ,
- (2) if  $R_t = \{z \in S \text{ s.t. } \psi_t(z) \neq \psi(z)\}$ , then  $\lambda^k(R_t) \rightarrow 0$ ,
- (3)  $\psi_t$  converges uniformly to  $\psi$ ,
- (4) the sequence  $\{\nabla\psi_t\}$  is a Cauchy sequence in  $L^p(S)$ .

Moreover, there exists a  $g \in L^p(S)$  such that for all  $t$ ,  $|\nabla\psi_t| \leq g$   $\lambda^k$ -a.e. on  $S$ , and the following is valid for all  $t$ :

$$\int_{S \cap B_2(0)} |\nabla\psi_t|^p \leq C(n), \quad \int_{S \cap B_1(0)} |\nabla\psi_t|^p \leq 1 + \epsilon. \quad (6.7)$$

*Proof.* The proof is a standard consequence of the properties proved so far for the map  $\phi$  and the usual Lusin-type approximation for  $W^{1,p}$  functions (see for example [EG92, theorem 3, sec 6.6.3]).

First of all, we fix some notation. Given the maps  $\phi_i : T \rightarrow T_i$  and  $\phi_\infty : T \rightarrow S$ , we set

$$\psi_i \equiv \phi_i^{-1} : T_i \rightarrow T, \quad \alpha_i \equiv \phi_\infty \circ \psi_i : T_i \rightarrow S, \quad \beta_i \equiv \alpha_i^{-1} : S \rightarrow T_i. \quad (6.8)$$

Note that, by the Reifenberg construction, all these maps are Hölder continuous maps.

Next, we introduce a slight variant of the function  $f(x)$  defined before. In particular, set for  $z \in S$ :

$$g(z) = \prod_{i=0}^{\infty} \sup_{y \in B_{3r_i}(\beta_i(z)) \cap T_i} |\nabla\sigma_i^{-1}|. \quad (6.9)$$

By adapting the proof of the  $W^{1,p}$  estimates in (5.99), we prove that  $g \in L^p(S)$  with

$$\int_{S \cap B_2(0)} g(z)^p d\lambda^k(z) \leq \int_{T \cap B_{2,1}(0)} g(\phi(x))^p J(\phi_\infty)|_x d\lambda^k(x) \leq 3C_1(n). \quad (6.10)$$

Moreover, we can also refine the bound to

$$\int_{S \cap B_1(0)} g(z)^p d\lambda^k(z) = \int_{T \cap \psi(S \cap B_1(0))} g(\phi(x))^p J(\phi_\infty)|_x d\lambda^k(x) \leq \omega_k(1 + \epsilon). \quad (6.11)$$

Now define for  $t \geq 1$  the sets  $R_t = g^{-1}[0, t]$ . Since  $g \in L^p$ , then

$$\limsup_{t \rightarrow \infty} t^p \lambda^k(R_t) \leq \limsup_{t \rightarrow \infty} \int_{g \geq t} g^p \rightarrow 0. \quad (6.12)$$

Moreover,  $\psi|_{R_t}$  is a Lipschitz function with Lipschitz constant bounded by  $c(n)t$ . Indeed, let  $x, y \in R_t$ , and set  $i$  to be such that  $r_{i+1}/2 < |x - y| \leq r_i/2$ . We prove that

$$|\psi(x) - \psi(y)| \leq c(n)t|x - y| \quad (6.13)$$

that is,  $\psi|_{R_t}$  is  $c(n)t$  Lipschitz wrt the extrinsic distance on  $S$ .

By definition,  $\psi(z) = \psi_i(\beta_i(z))$ . Moreover, the definition of  $g$  immediately implies a uniform Lipschitz condition on  $\psi_i$ . In particular,

$$\|\nabla\psi_i\|_{L^\infty(B_{r_i}(\beta_i(x)))} \leq \prod_{s=0}^i \|\nabla\sigma_s^{-1}\|_{L^\infty(B_{r_s}(\beta_s(x)))} \leq t. \quad (6.14)$$

Now, by (5.25), we have that  $|\beta_i(x) - x| + |\beta_i(y) - y| \leq c(n, \rho)\delta r_{i+1}$ , so, for  $\delta \leq \delta_0(n, \rho)$  small enough, we get

$$|\beta_i(x) - \beta_i(y)| \leq |x - y| + 2c\delta r_{i+1} \leq 2|x - y|, \quad (6.15)$$

and thus

$$|\psi(x) - \psi(y)| \leq |\psi_i(\beta_i(x)) - \psi_i(\beta_i(y))| \leq t|\beta_i(x) - \beta_i(y)| \leq 2t|x - y|, \quad (6.16)$$

as desired.

Now, define  $\psi_t$  to be a Lipschitz extension of  $\psi|_{R_t}$  over all  $\mathbb{R}^n$  with the same Lipschitz constant. In particular,  $\psi_t$  will be a Lipschitz function on  $S$  with Lipschitz constant bounded by  $2t$ . Note that, independently of  $t$ , we have the estimate

$$|\nabla\psi_t(z)| \leq 4g(z) \quad \text{for } \lambda^k - \text{a.e. } z \in S. \quad (6.17)$$

Indeed, if  $z \in R_t^C$ , this estimate is valid for all  $x$  for which  $\nabla\psi_t(x)$  is defined. Moreover, we prove that this estimate is also valid for all  $z \in R_t \subset S$  which have  $\lambda^k$ -density 1 (both with respect to  $R_t$  and  $S$ ) and such that  $\nabla\psi_t(x)$  is defined. Since  $R_t \subset S$  is  $k$ -rectifiable, it is well-known that the density of both these sets is 1  $\lambda^k$  almost everywhere (see for example [AFP00, theorem 2.63]). In particular, for  $\lambda^k$  a.e.  $z \in R_t$  we have

$$\lim_{r \rightarrow 0} \frac{\lambda^k(B_r(z) \cap R_t)}{\lambda^k(B_r(z) \cap S)} = 1. \quad (6.18)$$

For  $y \in S$ , set  $y_t$  to be an almost projection of  $y$  onto  $R_t$ , i.e., a point such that  $d(y, y_t) \leq 2d(y, R_t)$ . Then we have

$$\limsup_{y \rightarrow z} \frac{|\psi_t(y) - \psi_t(z)|}{|y - z|} \leq \limsup_{y \rightarrow z} \left[ \frac{|\psi_t(y) - \psi_t(y_t)|}{|y - z|} + \frac{|\psi_t(y_t) - \psi_t(z)|}{|y - z|} \right]. \quad (6.19)$$

By the proof of (6.13), and since  $|y_t - z| \leq 2|y - z|$ , we get

$$\limsup_{y \rightarrow z} \frac{|\psi_t(y_t) - \psi_t(z)|}{|y - z|} = \limsup_{y \rightarrow z} \frac{|\psi_t(y_t) - \psi_t(z)|}{|y - z|} \leq 4g(z). \quad (6.20)$$

Moreover, since  $z$  is a density 1 point on  $R_t \subset S$ ,  $|y - y_t|$  cannot be too big around  $z$ . Indeed, if  $|y - y_t| \geq 2r$ , then  $B_r(y) \cap R_t = \emptyset$ , and so

$$\left( \frac{1}{4} \limsup_{y \rightarrow z} \frac{|y - y_t|}{|y - z|} \right)^k = \limsup_{y \rightarrow z} \frac{\lambda^k(B_{|y-y_t|/2}(z) \cap S)}{\lambda^k(B_{2|y-z|}(z) \cap S)} \leq \limsup_{y \rightarrow z} \left( 1 - \frac{\lambda^k(B_{2|y-z|}(z) \cap R_t)}{\lambda^k(B_{2|y-z|}(z) \cap S)} \right) = 0. \quad (6.21)$$

By the global Lipschitz estimate on  $\psi_t$ , we get

$$\limsup_{y \rightarrow z} \frac{|\psi_t(y) - \psi_t(y_t)|}{|y - z|} \leq 2t \limsup_{y \rightarrow z} \frac{|y - y_t|}{|y - z|} = 0. \quad (6.22)$$



This proves (6.17). Now consider any  $T \geq t$ . Since  $S$  is rectifiable, and since  $\psi_t = \psi_T$  on  $R_t$ , then  $\nabla\psi_t = \nabla\psi_T$   $\lambda^k$ -a.e. on  $R_t$ . Moreover, by (6.17), we have

$$\int_S |\nabla\psi_T - \nabla\psi_t|^p = \int_{R_t^c} |\nabla\psi_T - \nabla\psi_t|^p \leq 2^p \int_{R_t^c} |\nabla\psi_t|^p + 2^p \int_{R_t^c} |\nabla\psi_T|^p \leq 2^p t^p \lambda^k(R_t^c) + 2^p \int_{R_t^c} |g|^p. \quad (6.23)$$

By (6.12), this proves point (4). Since  $\psi$  is a uniformly continuous function, point (3) is a corollary of point (2).  $\square$

**6.2. Proof of Theorem 3.4: The Discrete Rectifiable-Reifenberg.** Up to minor differences, the proof of this theorem is essentially the same as the proof of (5.73), and is actually much simpler from the technical point of view. Indeed, we do not need to define the sets  $\tilde{E}, \tilde{B}, \tilde{V}$ , but we can just deal with the sets  $E, V, B$ .

In particular, let  $r_\alpha = 2^{-\alpha}$  and  $\bar{r} = r_A$  for some  $A \in \mathbb{N}$ . One can define the measure

$$\mu \equiv \sum_{x_j \in S} \omega_k r_j^k \delta_{x_j}, \quad \mu_{\bar{r}} = \omega_k \sum_{x_j \in S \cap B_1(0) \text{ s.t. } r_j \geq \bar{r}} r_j^k \delta_{x_j}, \quad (6.24)$$

and prove inductively on  $\alpha = A, A-1, \dots, 0$  that for all  $x \in B_1(0)$  such that  $B_{r_\alpha/10}(x)$  is not contained in any of the  $B_{r_j}(x_j)$ :

$$\mu_{\bar{r}}(B_{r_\alpha}(x)) \leq D(n) r_\alpha^k. \quad (6.25)$$

By proceeding with a rough covering argument as in the proof of (5.13), one obtains easily a rough upper bound on  $\mu_{\bar{r}}(B_{r_{\alpha-1}}(x))$ . Moreover, by mimicking the inductive constructions in Section 5.4, one can define Excess sets, good, bad and final balls, and the maps  $\sigma_i$  and the approximating manifolds  $T_i$ . By studying the bi-Lipschitz properties of  $\sigma_i$ , and by keeping track of the holes created by final and bad balls in the same way as before, one proves the inductive estimate

$$\mu_{\bar{r}}(B_{r_\alpha}(x)) \leq C(n) \left( 1 + \sum_{r_\beta \leq 2r_\alpha} \int_{B_{r_\alpha}(x)} D_\mu^k(x, r_\alpha) d\mu(x) \right), \quad (6.26)$$

as desired.

## 7. $L^2$ -BEST APPROXIMATION THEOREMS

In this Section we prove the main estimate necessary for us to be able to apply the rectifiable-Reifenberg of Theorems 3.3 and 3.4 to the singular sets  $S_\epsilon^k(f)$  of the stratification induced for stationary harmonic maps.

Namely, we need to understand how to estimate on a ball  $B_r(x)$  the  $L^2$ -distance of  $S_\epsilon^k$  from the best approximating  $k$ -dimensional subspace. When  $r < \text{inj}(M)$  this means we would like to consider subspaces  $L^k \subseteq T_x M$  and estimate  $d^2(x, L^k)$  for  $x \in S^k \cap B_r(x)$ , where the distance is taken in the normal coordinate charts. After rescaling this is equivalent to looking at a ball of definite size, but assuming  $K_M \ll 1$ . The

main Theorem of this Section is stated in some generality as we will need apply it with some care when proving Theorems 1.3 and 1.4. We recall that by definition

$$W_\alpha(x) \equiv W_{r_\alpha, r_{\alpha-3}}(x) \equiv \theta_{r_{\alpha-3}}(x) - \theta_{r_\alpha}(x) \geq 0, \quad (7.1)$$

where  $r_\alpha = 2^{-\alpha}$ .

**Theorem 7.1** ( $L^2$ -Best Approximation Theorem). *Let  $f : B_{9r}(p) \rightarrow N$  be a stationary harmonic map satisfying (1.16) with  $r^{2-n} \int_{B_{9r}(p)} |\nabla f|^2 \leq \Lambda$ , and let  $\epsilon > 0$ . Then there exists  $\delta(n, K_N, \Lambda, \epsilon)$ ,  $C(n, K_N, \Lambda, \epsilon) > 0$  such that if  $K_M < \delta$ , and  $B_{9r}(p)$  is  $(0, \delta)$ -symmetric but not  $(k+1, \epsilon)$ -symmetric, then for any finite measure  $\mu$  on  $B_r(p)$  we have that*

$$D_\mu(p, r) \leq \inf_{L^k} r^{-2-k} \int_{B_r(p)} d^2(x, L^k) d\mu(x) \leq Cr^{-k} \int_{B_r(p)} W_0(x) d\mu(x), \quad (7.2)$$

where the inf is taken over all  $k$ -dimensional affine subspaces  $L^k \subseteq T_p M$ .

*Remark 7.1.* Throughout this section, for convenience we will fix  $r = 1$  without essential loss of generality.

*Remark 7.2.* The assumption  $K_M < \delta$  is of little consequence, since given any  $K_M$  this just means focusing the estimates on balls of sufficiently small radius after rescaling.

**7.1. Symmetry and Gradient Bounds.** In this subsection we study stationary harmonic maps which are not  $(k+1, \epsilon)$ -symmetric on some ball. In particular, we show that this forces the gradient to be of some definite size when restricted to any  $k+1$ -dimensional subspace. More precisely:

**Lemma 7.2.** *Let  $f : B_4(p) \subseteq M \rightarrow N$  be a stationary harmonic map satisfying (1.16) with  $\int_{B_4(p)} |\nabla f|^2 \leq \Lambda$ . Then for each  $\epsilon > 0$  there exists  $\delta(n, K_N, \Lambda, \epsilon) > 0$  such that if  $K_M < \delta$ , and  $B_4(p)$  is  $(0, \delta)$ -symmetric but is not  $(k+1, \epsilon)$ -symmetric, then for every  $k+1$ -subspace  $V^{k+1} \subseteq T_p M$  we have*

$$\int_{A_{3,4}(p)} |\langle \nabla f, V \rangle|^2 d\lambda \geq \delta, \quad (7.3)$$

where  $|\langle \nabla u, V \rangle|^2 \equiv \sum_1^{k+1} |\langle \nabla u, v_i \rangle|^2$  for any orthonormal basis  $\{v_i\}$  of  $V$ , and  $\lambda$  is the volume measure on  $M$  (which is equivalent to  $\lambda^n$ ).

*Proof.* The proof is by contradiction. So with  $n, K_N, \Lambda, \epsilon > 0$  fixed let us assume the result fails. Then there exists a sequence  $f_i : B_4(p_i) \rightarrow N_i$  for which  $K_{M_i} < \delta_i$ , with  $B_4(p_i)$  being  $(0, \delta_i)$ -symmetric but not  $(k+1, \epsilon)$ -symmetric, and such that for some subspaces  $V_i^{k+1}$  we have that

$$\int_{A_{3,4}(p_i)} |\langle \nabla f_i, V_i^{k+1} \rangle|^2 \leq \delta_i \rightarrow 0. \quad (7.4)$$

Then after passing to subsequences we have that  $V_i \rightarrow V^k \subseteq \mathbb{R}^n$  with

$$N_i \xrightarrow{C^{1,\alpha}} N, \quad (7.5)$$

and

$$f_i \rightarrow f : B_4(0^n) \rightarrow N, \quad (7.6)$$

where the convergence is in  $L^2 \cap H_{weak}^1$ . Now (7.4) and the  $H_{weak}^1$  convergence guarantees that

$$\int_{A_{3,4}(0^n)} |\langle \nabla f, v^{k+1} \rangle|^2 = 0. \quad (7.7)$$

On the other hand, the  $(0, \delta_i)$ -symmetry of the  $f_i$ , combined with the  $L^2$  convergence, tells us that  $f$  is 0-symmetric. Combining these tells us that

$$\int_{B_4(0^n)} |\langle \nabla f, v^{k+1} \rangle|^2 = 0, \quad (7.8)$$

and hence we have that  $f$  is  $k+1$ -symmetric. Because the convergence  $f_i \rightarrow f$  is in  $L^2$ , this contradicts that the  $f_i$  are not  $(k+1, \epsilon)$ -symmetric for  $i$  sufficiently large, which proves the Lemma.  $\square$

**7.2. Best  $L^2$ -Subspace Equations.** In order to prove Theorem 7.1 we need to identify which subspace minimizes the  $L^2$ -energy, and the properties about this subspace that allow us to estimate the distance. We begin in Section 7.2.1 by studying some very general properties of the second directional moments of a general probability measure  $\mu \subseteq B_1(p)$ . We will then study in Section 7.2.2 the Dirichlet energy of a stationary harmonic map when restricted to these directions.

**7.2.1. Second Directional Moments of a Measure.** Let us consider a probability measure  $\mu \subseteq B_1(0^n)$ , and let

$$x_{cm}^i = x_{cm}^i(\mu) \equiv \int x^i d\mu(x), \quad (7.9)$$

be the center of mass. Let us inductively consider the maximum of the second directional moments of  $\mu$ . More precisely:

**Definition 7.3.** Let  $\lambda_1 = \lambda_1(\mu) \equiv \max_{|v|^2=1} \int |\langle x - x_{cm}, v \rangle|^2 d\mu(x)$  and let  $v_1 = v_1(\mu)$  with  $|v_1| = 1$  (any of) the vector obtaining this maximum. Now let us define inductively the pair  $(\lambda_{k+1}, v_{k+1})$  from  $v_1, \dots, v_k$  by

$$\lambda_{k+1} = \lambda_{k+1}(\mu) \equiv \max_{|v|^2=1, \langle v, v_i \rangle=0} \int |\langle x - x_{cm}, v \rangle|^2 d\mu(x), \quad (7.10)$$

where  $v_{k+1}$  is (any of) the vector obtaining this maximum.

Thus  $v_1, \dots, v_n$  defines an orthonormal basis of  $\mathbb{R}^n$ , ordered so that they maximize the second directional moments of  $\mu$ . Let us define the subspaces

$$V^k = V^k(\mu) \equiv x_{cm} + \text{span}\{v_1, \dots, v_k\}. \quad (7.11)$$

The following is a simple but important exercise:

**Lemma 7.4.** *If  $\mu$  is a probability measure in  $B_1(0^n)$ , then for each  $k$  the functional*

$$\min_{L^k \subseteq \mathbb{R}^n} \int d^2(x, L^k) d\mu(x), \quad (7.12)$$

where the min is taken over all  $k$ -dimensional affine subspaces, attains its minimum at  $V^k$ . Further, we have that

$$\min_{L^k \subseteq \mathbb{R}^n} \int d^2(x, L^k) d\mu(x) = \int d^2(x, V^k) d\mu(x) = \lambda_{k+1}(\mu) + \cdots + \lambda_n(\mu). \quad (7.13)$$

Note that the best affine subspace  $V^k$  will necessarily pass through the center of mass  $x_{cm}$  by Steiner's formula, or equivalently by Jensen's inequality and the definition of  $x_{cm}$ .

Now let us record the following Euler-Lagrange formula, which is also an easy computation:

**Lemma 7.5.** *If  $\mu$  is a probability measure in  $B_1(0^n)$ , then we have that  $v_1(\mu), \dots, v_n(\mu)$  satisfy the Euler-Lagrange equations:*

$$\int \langle x - x_{cm}, v_k \rangle (x - x_{cm})^i d\mu(x) = \lambda_k v_k^i, \quad (7.14)$$

where

$$\lambda_k = \int |\langle x - x_{cm}, v_k \rangle|^2 d\mu(x). \quad (7.15)$$

*Proof.* The proof is a simple application of the Lagrange multipliers method. Inductively on  $k$ , consider the function  $f(v_k, \lambda_k, \lambda_{k,1}, \dots, \lambda_{k,k-1}) : \mathbb{R}^n \times \mathbb{R}^k \rightarrow \mathbb{R}$  given by

$$f(v_k, \lambda_k, \lambda_{k,\ell}) = \int \langle x - x_{cm}, v_k \rangle^2 d\mu(x) - \lambda_k (|v_k|^2 - 1) - 2 \sum_{\ell=0}^{k-1} \lambda_{k,\ell} \langle v_k, v_\ell \rangle. \quad (7.16)$$

By the multipliers method, we obtain that  $v_k$  and  $\lambda_k$  satisfy the equations

$$\frac{1}{2} \nabla^{(v_k)} f(v_k, \lambda_k, \lambda_{k,\ell}) = \int \langle x - x_{cm}, v_k \rangle (x - x_{cm}) d\mu(x) - \lambda_k v_k - \sum_{\ell=0}^{k-1} \lambda_{k,\ell} v_\ell = 0. \quad (7.17)$$

By taking the scalar product of this equation with  $v_\ell$ , since  $\langle v_k, v_\ell \rangle = \delta_{k,\ell}$ , we have

$$\lambda_k = \int |\langle x - x_{cm}, v_k \rangle|^2 d\mu(x), \quad (7.18)$$

$$\begin{aligned} \lambda_{k,\ell} &= \int \langle x - x_{cm}, v_\ell \rangle \langle x - x_{cm}, v_k \rangle d\mu(x) = \left\langle \int \langle x - x_{cm}, v_\ell \rangle (x - x_{cm}) d\mu(x), v_k \right\rangle = \\ &= \left\langle \lambda_\ell v_\ell + \sum_{s=1}^{\ell-1} \lambda_{\ell,s} v_s, v_k \right\rangle = 0. \end{aligned} \quad (7.19)$$

□

7.2.2. *Restricted Dirichlet Energies.* Our goal is to now study the Dirichlet energy of a stationary harmonic map when restricted to the directions  $v_1(\mu), \dots, v_n(\mu)$  associated to a probability measure. The main result of this subsection is the following, which holds for a general stationary harmonic map:

**Proposition 7.6.** *Let  $f : B_9(p) \subseteq M \rightarrow N$  be a stationary harmonic map satisfying (1.16) with  $K_M < 2^{-4}$  and such that  $\int_{B_9(p)} |\nabla f|^2 \leq \Lambda$ . Let  $\mu$  be a probability measure on  $B_1(p)$  with  $\lambda_k(\mu), v_k(\mu)$  defined as in Definition 7.3. Then there exists  $C(n, K_N) > 0$  such that*

$$\lambda_k \int_{A_{3,4}(p)} |\langle \nabla f(z), v_k \rangle|^2 dv_g(z) \leq C \int W_0(x) d\mu(x). \quad (7.20)$$

*Proof.* Note first that there is no harm in assuming that  $x_{cm} \equiv 0$ . If not we can easily translate to make this so, in which case we still have that  $\text{supp}(\mu) \subseteq B_2$ . Additionally, we will simplify the technical aspect of the proof by assuming that  $M \equiv \mathbb{R}^n$ . By working in normal coordinates the proof of the general case is no different except up to some mild technical work.

Now let us fix any  $z \in A_{3,4}$ . By Lemma 7.5, we can inner product both sides of (7.14) by  $\nabla f(z)$  to obtain for each  $k$  and  $z \in A_{3,4}$ :

$$\lambda_k \langle \nabla f(z), v_k \rangle = \int \langle x, v_k \rangle \langle \nabla f(z), x \rangle d\mu(x), \quad (7.21)$$

Observe that, by definition of center of mass,

$$\int \langle x, z \rangle d\mu(x) = \langle x_{cm}, z \rangle = 0. \quad (7.22)$$

Thus we can write

$$\lambda_k \langle \nabla f(z), v_k \rangle = \int \langle x, v_k \rangle \langle \nabla f(z), x - z \rangle d\mu(x). \quad (7.23)$$

By Hölder inequality, we can then estimate

$$\lambda_k^2 |\langle \nabla f(z), v_k \rangle|^2 \leq \lambda_k \int |\langle \nabla f(z), x - z \rangle|^2 d\mu(x). \quad (7.24)$$

If  $\lambda_k \neq 0$ , integrating with respect to  $z$  on both sides we get the estimate

$$\lambda_k \int_{A_{3,4}} |\langle \nabla f(z), v_k \rangle|^2 dv_g(z) \leq \int \int_{A_{3,4}} |\langle \nabla f(z), x - z \rangle|^2 dv_g(z) d\mu(x). \quad (7.25)$$

Set for convenience  $n_x(z) = (z - x) / |z - x|$ , i.e.,  $n_x(z)$  is the radial vector from  $x$  to  $z$ . Now for  $x \in \text{supp}(\mu)$  we can estimate

$$\begin{aligned} \int_{A_{3,4}} |\langle \nabla f(z), x - z \rangle|^2 dv_g(z) &= \int_{A_{3,4}} |\langle \nabla f(z), n_x(z) \rangle|^2 |x - z|^{2-n} |x - z|^n dv_g(z) \\ &\leq 6^n \int_{A_{3,4}} |\langle \nabla f(z), n_x(z) \rangle|^2 |x - z|^{2-n} dv_g(z) \\ &\leq C(n) \int_{A_{1,8}(x)} |\langle \nabla f(z), n_x(z) \rangle|^2 |x - z|^{2-n} dv_g(z) \\ &= C(n) W_0(x). \end{aligned} \quad (7.26)$$

Applying this to (7.25) we get as desired

$$\lambda_k \int_{A_{3,4}} |\langle \nabla f(z), v_k \rangle|^2 dv_g(z) \leq C(n) \int W_0(x) d\mu(x). \quad (7.27)$$

□

**7.3. Proof of Theorem 7.1.** Let us now combine the results of this Section in order to prove Theorem 7.1. Indeed, let  $\mu$  be a measure in  $B_1(p) \subseteq T_p M$ . We can assume that  $\mu$  is a probability measure without any loss of generality, since both sides of our estimate scale. Let  $(\lambda_1(\mu), v_1(\mu)), \dots, (\lambda_n(\mu), v_n(\mu))$  be the directional second moments as defined in Definition 7.3, with  $V^k$  the induced subspaces defined as in (7.11). Using Lemma 7.4 we have that

$$\min_{L^k \subseteq \mathbb{R}^n} \int d^2(x, L^k) d\mu(x) = \int d^2(x, V^k) d\mu(x) = \lambda_{k+1}(\mu) + \dots + \lambda_n(\mu) \leq (n-k)\lambda_{k+1}(\mu), \quad (7.28)$$

where we have used that  $\lambda_j \leq \lambda_i$  for  $j \geq i$ . Therefore our goal is to estimate  $\lambda_{k+1}$ . To begin with, Proposition 7.6 tells us that for each  $j$

$$\lambda_j \int_{A_{3,4}(p)} |\langle \nabla f(z), v_j \rangle|^2 dv_g(z) \leq C \int W_0(x) d\mu(x). \quad (7.29)$$

Let us sum the above for all  $j \leq k+1$  in order to obtain

$$\sum_{j=1}^{k+1} \lambda_j \int_{A_{3,4}(p)} |\langle \nabla f(z), v_j \rangle|^2 dv_g(z) \leq (k+1)C \int W_0(x) d\mu(x), \quad (7.30)$$

or by using that  $\lambda_{k+1} \leq \lambda_j$  for  $k+1 \geq j$  we get

$$\lambda_{k+1} \int_{A_{3,4}(p)} |\langle \nabla f(z), V^{k+1} \rangle|^2 dv_g(z) = \lambda_{k+1} \sum_{j=1}^{k+1} \int_{A_{3,4}(p)} |\langle \nabla f(z), v_j \rangle|^2 dv_g(z) \leq C \int W_0(x) d\mu(x). \quad (7.31)$$

Now we use that  $B_8(p)$  is  $(0, \delta)$ -symmetric, but *not*  $(k+1, \epsilon)$ -symmetric in order to apply Lemma 7.2 and conclude that

$$\int_{A_{3,4}(p)} |\langle \nabla f(z), V^{k+1} \rangle|^2 dv_g(z) \geq \delta. \quad (7.32)$$

Combining this with (7.31) we obtain

$$\delta \lambda_{k+1} \leq \lambda_{k+1} \int_{A_{3,4}(p)} |\langle \nabla f(z), V^{k+1} \rangle|^2 dv_g(z) \leq C \int W_0(x) d\mu(x), \quad (7.33)$$

or that

$$\lambda_{k+1} \leq C(n, K_N, \Lambda, \epsilon) \int W_0(x) d\mu(x). \quad (7.34)$$

Combining this with (7.28) we have therefore proved the Theorem. □

## 8. THE INDUCTIVE COVERING LEMMA

This Section is dedicated to the basic covering lemma needed for the proof of the main theorems of the paper. The covering scheme is similar in nature to the one introduced by the authors in [NV] in order to prove structural theorems on critical sets. Specifically, let us consider a stationary harmonic map  $f$  between Riemannian manifolds. We wish to build a covering of the quantitative stratification

$$S_{\epsilon,r}^k(f) \cap B_1(p) \subseteq U_r \cup U_+ = U_r \cup \bigcup B_{r_i}(x_i), \quad (8.1)$$

which satisfies several basic properties. To begin with, the set  $U_+$  is a union of balls satisfying  $r_i > r \geq 0$ , and should satisfy the packing estimate  $\omega_k \sum r_i^k \leq C$ . Each ball  $B_{r_i}(x_i)$  should have the additional property that there is a definite energy drop of  $f$  when compared to  $B_1(p)$ . To describe the set  $U_r$  we should distinguish between the case  $r > 0$  and  $r \equiv 0$ . In the case  $r > 0$  we will have that  $U_r = \bigcup B_r(x'_i)$  is a union of  $r$ -balls and satisfies the Minkowski estimate  $\text{Vol}(U_r) \leq Cr^{n-k}$ . In the case when  $r \equiv 0$  we will have that  $U_0$  is  $k$ -rectifiable with the Hausdorff estimate  $\lambda^k(U_0) \leq C$ . Let us be more precise:

**Lemma 8.1** (Covering Lemma). *Let  $f : B_{16}(p) \subseteq M \rightarrow N$  be a stationary harmonic map satisfying (1.16) with  $\int_{B_{16}(p)} |\nabla f|^2 \leq \Lambda$  and  $K_M < \delta(n, K_N, \Lambda, \epsilon)$ . Let  $E = \sup_{x \in B_1(p) \cap S_{\epsilon,r}^k} \theta_1(x)$  with  $\epsilon > 0$ ,  $r \geq 0$ , and  $k \in \mathbb{N}$ . Then for all  $\eta \leq \eta(n, K_N, \Lambda, \epsilon)$ , there exists a covering  $S_{\epsilon,r}^k(f) \cap B_1(p) \subseteq U = U_r \cup U_+$  such that*

- (1)  $U_+ = \bigcup B_{r_i}(x_i)$  with  $r_i > r$  and  $\sum r_i^k \leq C(n, K_N, \Lambda, \epsilon)$ .
  - (2)  $\sup_{y \in B_{r_i}(x_i) \cap S_{\epsilon,r}^k} \theta_{r_i}(y) \leq E - \eta$ .
  - (3) If  $r > 0$  then  $U_r = \bigcup_1^N B_r(x'_i)$  with  $N \leq C(n)r^{-k}$ .
  - (4) If  $r = 0$  then  $U_0$  is  $k$ -rectifiable and satisfies  $\text{Vol}(B_s(U_0)) \leq C(n)s^{n-k}$  for each  $s > 0$ .
- In particular,  $\lambda^k(U_0) \leq C(n)$ .

*Remark 8.1.* The assumption  $K_M < \delta$  is of little consequence, since given any  $K_M$  this just means focusing the estimates on balls of sufficiently small radius after rescaling.

To prove the result let us begin by outlining the construction of the covering, we will then spend the rest of this section proving the constructed cover has all the desired properties.

Thus let us consider some  $\eta > 0$  fixed, and then define the energy scale for  $x \in B_1(p)$  by

$$s_x = s_x^{E,\eta} \equiv \inf \{r \leq s \leq 1 : \sup_{B_s(x) \cap S_{\epsilon,r}^k} \theta_{\eta s}(y) \geq E - \eta\}. \quad (8.2)$$

Note that the energy scale implicitly depends on many constants. If  $r = 0$  let us define the set  $U_0$  by

$$U_0 \equiv \{x \in S_{\epsilon,r}^k(f) \cap B_1(p) : s_x = 0\}, \quad (8.3)$$

while if  $r > 0$  let us first define the temporary covering

$$U'_r = \bigcup B_r(x'_i), \quad (8.4)$$

where

$$\{x'_i\} \subseteq \{x \in S_{\epsilon,r}^k(f) \cap B_1(p) : s_x = r\}, \quad (8.5)$$

is any minimal  $r/5$ -dense set. In particular, note that the collection of balls  $\{B_{r/10}(x_i^r)\}$  are disjoint.

In order to define the temporary covering  $U'_+ = \{B_{r_i}(x_i)\}$  let us first consider

$$\{x \in S_{\epsilon,r}^k(f) \cap B_1(p) : s_x > r\} \subseteq \bigcup_{s_x > r} B_{s_x/10}(x), \quad (8.6)$$

and choose from it a Vitali subcovering so that

$$\{x \in S_{\epsilon,r}^k(f) \cap B_1(p) : s_x > r\} \subseteq \bigcup_{i \in I} B_{r_i/2}(x_i) \subseteq \bigcup_{i \in I} B_{r_i}(x_i) \equiv U'_+, \quad (8.7)$$

where  $r_i \equiv s_{x_i}$ . In particular, we have that the collection of balls  $\{B_{r_i/10}(x_i)\}$  are all disjoint. It is clear from the construction that we have built a covering

$$S_{\epsilon,r}^k(f) \cap B_1(p) \subseteq \bigcup B_{r/2}(x_i^r) \cup \bigcup_{i \in I} B_{r_i/2}(x_i) \subseteq \bigcup B_r(x_i^r) \cup \bigcup_{i \in I} B_{r_i}(x_i) \equiv U'_r \cup U'_+. \quad (8.8)$$

Now this is not quite the covering of Lemma 8.1, as the energy drop properties of (8.2) involve dropping an extra  $\eta$ -scales. It will be more convenient to estimate this covering, however with only minimal work let us now build from this covering the desired covering of Lemma 8.1, which is only a small modification. Indeed, consider for each ball  $B_{r_i}(x_i)$  (or  $B_r(x_i^r)$ ) a Vitali covering

$$B_{r_i/2}(x_i) \cap S_{\epsilon,r}^k \subseteq \bigcup_1^{N_i} B_{\eta r_i}(x_{ia}) \equiv \bigcup_1^{N_i} B_{r_{ia}}(x_{ia}), \quad (8.9)$$

where  $x_{ia} \in B_{r_i/2}(x_i) \cap S_{\epsilon,r}^k$  and by a standard covering argument  $N_i \leq N_i(n, \eta)$ . Then we can define the coverings

$$\begin{aligned} U_r &\equiv \bigcup_i \bigcup_{a=1}^{N_i} B_{r_{ia}}(x_{ia}^r), \\ U_+ &\equiv \bigcup_i \bigcup_{a=1}^{N_i} B_{r_{ia}}(x_{ia}). \end{aligned} \quad (8.10)$$

It is clear from the construction that  $U_+$  now satisfies the energy drop condition of Lemma 8.1.2. What is left is to show the content estimates of Lemma 8.1, and from our estimates on  $N_i$  it is clear with  $\eta < \eta(n, K_N, \Lambda, \epsilon)$  that it is enough to prove these estimates for the sets  $U'_r$  and  $U'_+$  themselves, which is therefore the goal of much of this section.

The outline of this Section is as follows. Section 8.1 is dedicated to proving a variety of technical lemmas which will be used to further decompose the sets  $U'_+$  and  $U'_r$  when  $r > 0$ . The technical issue is that we cannot directly apply the discrete Reifenberg to the set  $U'_+$ , and will instead need to exchange  $U'_+$  for a more manageable collection of balls without losing much content. In Section 8.2 we will use these tools in order to prove our content estimates on  $U'_+$  and  $U'_r$  when  $r > 0$ . The proof will require an inductive argument combined with applications of the discrete Reifenberg of Theorem 3.4 to our new covering. Finally in Section 8.3 we will prove the desired estimates on  $U_0$ . The volume estimates will follow almost immediately from our previous constructions, and to prove the rectifiable statement will require a careful application of



the rectifiable Reifenberg theorem.

**8.1. Technical Constructions for Decomposing  $U'_r \cup U'_+$ .** Let us consider the set of positive radius balls in our covering given by

$$U_{>} \equiv \begin{cases} U'_+ & \text{if } r = 0, \\ U'_r \cup U'_+ & \text{if } r > 0. \end{cases} \quad (8.11)$$

This section will be dedicated to proving some technical results which will be required in estimating this set. The estimates on this set are a little delicate, the reason being that we cannot directly apply the rectifiable Reifenberg of Theorem 3.4 to this set. Instead, we will need to replace  $U_{>}$  with a different covering at each stage, which will be more adaptable to Theorem 3.4. Thus this subsection is dedicated to proving a handful of technical results which are important in the construction of this new covering.

Throughout this subsection we are always working under the assumptions of Lemma 8.1. Let us begin with the following point, which is essentially a consequence of the continuity of the energy:

**Lemma 8.2.** *For each  $\eta' > 0$  there exists  $R(n, K_N, \Lambda, \eta') > 1$  such that if  $\delta < \delta(n, K_N, \Lambda, \eta')$  and  $\eta \leq \eta(n, K_N, \Lambda, \eta')$ , then we have for each  $z \in B_{r_i}(x_i)$  the estimate*

$$\theta_{Rr_i}(z) > E - \eta'. \quad (8.12)$$

*Proof.* The proof relies on a straight forward energy comparison. Namely, let  $x, y \in B_{s/2}$  with  $s < 1$  and let us denote  $d \equiv d(x, y)$ . Then we have the estimate

$$\theta_s(y) = s^{2-n} \int_{B_s(y)} |\nabla f|^2 \leq s^{2-n} \int_{B_{s+d}(x)} |\nabla f|^2 = \left( \frac{s}{s+d} \right)^{2-n} \theta_{s+d}(x). \quad (8.13)$$

To apply this in our context, let us note for each  $x_i$  in our covering, that by our construction of  $U_{>}$  there must exist  $y_i \in B_{r_i}(x_i)$  such that  $\theta_{(R-1)r_i}(y_i) \geq \theta_{r_i}(y_i) = E - \eta$ . Let us now apply (8.13) to obtain

$$\theta_{Rr_i}(z) \geq \left( \frac{Rr_i}{(R-2)r_i} \right)^{2-n} \theta_{(R-2)r_i}(y_i) \geq \left( \frac{R}{R-2} \right)^{2-n} (E - \eta). \quad (8.14)$$

If  $R = R(n, K_N, \Lambda, \eta') > 0$  and  $\eta \leq \eta(n, K_N, \Lambda, \eta')$ , then we obtain from this the claimed estimate.  $\square$

In words, the above Lemma is telling us that even though we have no reasonable control over the size of  $\theta_{r_i}(x_i)$ , after we go up a controlled number of scales we can again assume that the energy density is again close to  $E$ .

In the last Lemma the proof was based on continuity estimates on the energy  $\theta$ . In the next Lemma we wish to show an improved version of this continuity under appearance of symmetry. Precisely:

**Lemma 8.3** (Improved Continuity of  $\theta$ ). *Let  $f : B_4(p) \rightarrow N$  be a stationary harmonic map satisfying (1.16) with  $\int_{B_4(p)} |\nabla f|^2 \leq \Lambda$ . Then for each  $0 < \tau, \eta < 1$  there exists  $\delta(n, K_N, \Lambda, \eta, \tau) > 0$  such that if*

- (1) we have  $K_M < \delta$ ,

(2) there exists  $x_0, \dots, x_k \in B_1(p)$  which are  $\tau$ -independent with  $|\theta_\delta(x_j) - \theta_3(x_j)| < \delta$ ,  
then if  $V^k$  is the  $k$ -dimensional subspace spanned by  $x_0, \dots, x_k$ , then for all  $x, y \in B_1(p) \cap B_\delta(V^k)$  and  $\eta \leq s \leq 1$  we have that  $|\theta_s(x) - \theta_s(y)| < \eta$ .

*Proof.* The proof is by contradiction. Thus, imagine no such  $\delta$  exists. Then there exists a sequence of stationary harmonic maps  $f_i : B_4(p_i) \rightarrow N_i$  satisfying  $\int_{B_4(p)} |\nabla f|^2 \leq \Lambda$  such that

(1)  $K_M < \delta_i \rightarrow 0$

(2) there exists  $x_{i,0}, \dots, x_{i,k} \in B_1(p)$  which are  $\tau$ -independent with  $|\theta_{\delta_i}^{f_i}(x_{i,j}) - \theta_3^{f_i}(x_{i,j})| < \delta_i \rightarrow 0$ ,

however we have that there exists  $x_i, y_i \in B_1(p) \cap B_{\delta_i}(V_i^k)$  and  $\eta \leq s_i \leq 1$  such that  $|\theta_{s_i}^{f_i}(x_i) - \theta_{s_i}^{f_i}(y_i)| \geq \eta$ . Let us now pass to a subsequence to obtain the defect measure

$$\begin{aligned} |\nabla f_i|^2 dv_g &\rightarrow |\nabla f|^2 dv_g + \nu, \\ x_{i,\beta} &\rightarrow x_\beta \in B_1(0^n), \\ x_i &\rightarrow x, y_i \rightarrow y \in V = \text{span}\{x_0, \dots, x_k\}, \\ s_i &\rightarrow s, \end{aligned} \tag{8.15}$$

where  $\nu$  is a measure on  $B_4(0^n)$ , and

$$s^{2-n} \left| \int_{B_s(x)} (|\nabla f|^2 dv_g + d\nu) - \int_{B_s(y)} (|\nabla f|^2 dv_g + d\nu) \right| \geq \eta. \tag{8.16}$$

However, we have by theorem 2.7 that  $f$  and  $\nu$  are 0-symmetric on  $B_2$  with respect to each of the points  $x_0, \dots, x_k$ . In particular, by the standard cone splitting arguments we have that  $\nu$  is  $k$ -symmetric with respect to the  $k$ -plane  $V$  on  $B_1$ . In particular,  $\nu$  is invariant under translation by elements of  $V$ . However, this is a contradiction to (8.16), and therefore we have proved the result.  $\square$

Now the first goal is to partition  $U_>$  into a finite collection, each of which will have a few more manageable properties than  $U_>$  itself. More precisely:

**Lemma 8.4.** *For each  $R > 1$  there exists  $N(n, R) > 1$  such that we can break up  $U_>$  as a union*

$$U_> = \bigcup_{a=1}^N U_>^a = \bigcup_{a=1}^N \bigcup_{i \in I^a} B_{r_i}(x_i), \tag{8.17}$$

such that each  $U_>^a$  has the following property: if  $i \in I^a$ , then for any other  $j \in I^a$  we have that if  $x_j \in B_{Rr_i}(x_i)$ , then  $r_j < R^{-2}r_i$ .

*Proof.* Let us recall that the balls in the collection  $\{B_{r_i/5}(x_i)\}$  are pairwise disjoint. In particular, given  $R > 1$  if we fix a ball  $B_{r_i}(x_i)$  then by the usual covering arguments there can be at most  $\bar{N}(n, R)$  ball centers  $\{x_j\}_1^N \in U_> \cap B_{R^3r_i}(x_i)$  with the property that  $r_j \geq R^{-6}r_i$ . Indeed, if  $\{x_j\}_1^N$  is such a collection of balls then we get

$$\omega_n(6R^3r_i)^n = \text{Vol}(B_{6R^3r_i}(x_i)) \geq \sum_1^N \text{Vol}(B_{r_j/5}(x_j)) \geq N\omega_n(R^{-6}r_i/5)^n, \tag{8.18}$$

which by rearranging gives the estimate  $\bar{N} \leq \bar{N}(n, R)$ .

Now we wish to build our decomposition  $U_{>} = \bigcup_1^N U_{>}^a$ , where  $N = \bar{N} + 1$  is from the first paragraph. We shall do this inductively, with the property that at each step of the inductive construction we will have that the sets  $U_{>}^a$  will satisfy the desired property. In particular, for every  $a$  and  $i \in I^a$ , if  $j \in I^a$  is such that  $x_j \in B_{Rr_i}(x_i)$ , then  $r_j < R^{-2}r_i$ .

Begin by letting each  $I^a$  be empty. We are going to sort the points  $\{x_i\}_{i \in I}$  into the sets  $I^a$  one at a time. At each step let  $i \in I \setminus \bigcup_{a=1}^N I^a$  be an index such that  $r_i = \max r_j$ , where the max is taken over all indexes in  $I \setminus \bigcup_{a=1}^N I^a$ , i.e., over all indexes which haven't been sorted out yet. Now let us consider the collection of ball centers  $\{y_j\}_{j \in J}$  such that  $J \subset I$ ,  $x_i \in B_{Rr_j}(y_j)$  and  $r_i \leq r_j \leq R^2r_i$ . Note that, by construction, either  $y_j$  has already been sorted out in some  $I^a$ , or  $r_j = r_i$ . Now evidently  $y_j \in B_{R^3r_i}(x_i)$  for all  $j \in J$  and so by the first paragraph the cardinality of  $J$  is at most  $N(n, R)$ . In particular, there must be some  $I^a$  such that  $I^a \cap \{y_j\} = \emptyset$ . Let us assign  $i \in I^a$  to this piece of the decomposition. Clearly the decomposition  $\bigcup U_{>}^a$  still satisfies the inductive hypothesis after the addition of this point, and so this finishes the inductive step of construction. Since at each stage we have chosen  $x_i$  to have the maximum radius, this process will continue indefinitely to give the desired decomposition of  $U_{>}$ . □

Now with a decomposition fixed, let us consider for each  $1 \leq a \leq N$  the measures

$$\mu^a \equiv \sum_{x_i \in U_{>}^a} \omega_k r_i^k \delta_{x_i}. \quad (8.19)$$

The following is a crucial point in our construction. It tells us that each ball  $B_{10r_i}(x_i)$  either has small  $\mu^a$ -volume, or the point  $x_i$  must have large energy at scale  $r_i$ . Precisely:

**Lemma 8.5.** *Let  $D, \eta' > 0$  be fixed. There exists  $R = R(n, K_N, \Lambda, \eta', D, \epsilon) > 0$  such that if we consider the decomposition (8.17), and if*

- (1)  $K_M < \delta(n, K_N, \Lambda, \eta', \epsilon)$  and  $\eta \leq \eta(n, K_N, \Lambda, \eta', \epsilon)$ ,  $r_i < 10^{-4}$ ,
- (2) we have  $\mu^a(B_{10r_i}(x_i)) \geq 2\omega_k r_i^k$ ,
- (3) for all ball centers  $y_j \in A_{r_i/10, 10r_i}(x_i) \cap U^a$  and for  $s = 10^{-2n} D^{-1} r_i$ , we have that  $\mu^a(B_s(y_j)) \leq Ds^k$ ,

then we have that  $\theta(x_i, \eta' r_i) \geq E - \eta'$ .

*Proof.* Let us begin by choosing  $\eta'' \ll \eta'$ , which will be fixed later in the proof so that  $\eta'' = \eta''(n, K_N, \Lambda, \epsilon)$ , and let us also define  $\tau \equiv 10^{-3n} D^{-1}$ . Let  $\delta(n, K_N, \Lambda, \eta')$  be from Lemma 8.3 so that the conclusions hold with  $10^{-1}\eta'$ , and let  $\delta'(n, K_N, \Lambda, \eta'', \epsilon)$  be chosen so that conclusions of Lemma 8.2 are satisfied with  $\eta''$  if  $\eta \leq \eta(n, K_N, \Lambda, \eta'', \epsilon)$ . Now throughout we will assume  $\eta < \delta$  and  $K_M < \min\{\delta, \delta'\}$ . We will also choose  $R = R(n, K_N, \Lambda, \eta'', D, \epsilon) > \max\{\tau^{-1}, \delta^{-1}, \delta'^{-1}\}$  so that Lemma 8.2 is satisfied with  $\eta''$ .

Since it will be useful later, let us first observe that  $r_i \geq Rr$ . Indeed, if not then for each ball center  $y_j \in A_{r_i/10, 10r_i}(x_i) \cap U_{>}^a$  we would have  $r \leq r_j < R^{-2}r_i < r$ . This tells us that there can be no ball centers in

$A_{r_i/10,10r_i}(x_i) \cap U_{>}^a$ . However, by our volume assumption we have that

$$\mu^a(A_{r_i/10,10r_i}(x_i)) = \mu^a(B_{10r_i}(x_i)) - \mu^a(B_{r_i/10}(x_i)) \geq 2\omega_k r_i^k - \omega_k r_i^k \geq \omega_k r_i^k, \quad (8.20)$$

which contradicts this. Therefore we must have that  $r_i \geq Rr$ .

Now our first real claim is that under the assumptions of the Lemma there exists ball centers  $y_0, \dots, y_k \in U_{>}^a \cap A_{r_i/10,10r_i}(x_i)$  which are  $\tau r_i$ -independent in the sense of Definition 2.2. Indeed, assume this is not the case, then we can find a  $k-1$ -plane  $V^{k-1}$  such that

$$\{y_i\}_{i \in I^a} \cap A_{r_i/10,10r_i}(x_i) \subseteq B_{\tau r_i}(V^{k-1}). \quad (8.21)$$

In particular, by covering  $B_{\tau r_i}(V) \cap B_{10r_i}(x_i) \cap I^a$  by  $C(n)\tau^{1-k} \leq 10^n \tau^{1-k}$  balls of radius  $10\tau r_i$  with centers in  $I^a$ , and using our assumption that  $\mu^a(B_{\tau r_i}(y)) \leq D\tau^k r_i^k$ , we are then able to conclude the estimate

$$\mu^a(A_{r_i/10,10r_i}(x_i)) \leq \mu^a(B_{\tau r_i}(V) \cap B_{10r_i}(x_i)) \leq 10^{2n} D\tau r_i^k < \omega_n r_i^k. \quad (8.22)$$

On the other hand, our volume assumption guarantees that

$$\mu^a(A_{r_i/10,10r_i}(x_i)) = \mu^a(B_{10r_i}(x_i)) - \mu^a(B_{r_i/10}(x_i)) \geq 2\omega_k r_i^k - \omega_k r_i^k \geq \omega_k r_i^k, \quad (8.23)$$

which leads to a contradiction. Therefore there must exist  $k+1$  ball centers  $y_0, \dots, y_k \in A_{r_i/10,10r_i}(x_i)$  which are  $\tau r_i$ -independent points, as claimed.

Let us now remark on the main consequences of the existence of these  $k+1$  points. Note first that for each  $y_j$  we have that  $\theta_{R^{-1}r_i}(y_j) > E - \eta''$ , since by the construction of  $U_{>}^a$  we have that  $r_j \leq R^{-2}r_i$ , and therefore we can apply Lemma 8.2. Thus we have  $k+1$  points in  $B_{10r_i}(x_i)$  which are  $\tau r_i$ -independent, and whose energies are  $\eta''$ -pinched. To exploit this, let us first apply Lemma 8.3 in order to conclude that for each  $x \in B_\delta(V)$ , where  $V$  is the plane spanned by the  $k+1$  independent points just determined, we have

$$\theta_{\eta' r_i}(x) \geq \theta_{\eta' r_i}(y_j) - |\theta_{\eta' r_i}(x) - \theta_{\eta' r_i}(y_j)| \geq E - \eta'' - 10^{-1}\eta' > E - \eta'. \quad (8.24)$$

In particular, if we assume that  $x_i$  is such that  $\theta_{\eta' r_i}(x_i) < E - \eta'$ , then we must have that  $r_i^{-1}d(x_i, V) \geq \delta = \delta(n, K_N, \Lambda, \eta')$ .

Therefore, let us now assume  $\theta_{\eta' r_i}(x_i) < E - \eta'$ , and thus to prove the Lemma we wish to find a contradiction. To accomplish this notice that we have our  $k+1$  points  $y_0, \dots, y_k \in B_{10r_i}$  which are  $\tau r_i$ -independent and for which  $|\theta_{20r_i}(y_j) - \theta_{R^{-1}r_i}(y_j)| < \eta''$ . Therefore by applying the cone splitting of Theorem 2.3 we have for each  $\epsilon' > 0$  that if  $\eta'' \leq \eta''(n, K_N, \Lambda, \epsilon')$  then  $B_{10r_i}(x_i)$  is  $(k, \epsilon')$ -symmetric with respect to the  $k$ -plane  $V^k$ . However, since  $d(x_i, V) > \delta r_i$ , we have by Theorem 2.4 that if  $\epsilon' \leq \epsilon'(n, K_N, \Lambda, \epsilon)$  then there exists some  $\tau' = \tau'(n, K_N, \Lambda, \epsilon)$  such that  $B_{\tau' r_i}(x_i)$  is  $(k+1, \epsilon)$ -symmetric. However, we can assume after a further increase that  $R = R(n, K_N, \Lambda, D, \epsilon) > 4\tau'^{-1}$ , and thus we have that  $\tau' r_i > 4R^{-1}r_i > 4r$ . This contradicts that  $x_i \in S_{\epsilon, r}^k$ , and thus we have contradicted that  $\theta_{\eta' r_i}(x_i) < E - \eta'$ , which proves the Lemma.  $\square$

**8.2. Estimating  $U_>$  in Lemma 8.1.** Recall the set  $U_>$  defined in (8.11), which consists of all positive radii balls in our covering. We now wish to estimate this set in this subsection. First let us pick  $D' = D'(n) \equiv 2^{16n}D(n)$ , where  $D(n)$  is from Theorem 3.4. Then for some  $\eta'$  fixed we can choose  $R$  as in Lemma 8.5. It is then enough to estimate each of the sets  $U_>^a$ , as there are at most  $N = N(n, K_N, \Lambda, \epsilon, \eta, \eta')$  pieces to the decomposition. Thus we will fix a set  $U_>^a$  and focus on estimating the content of this set. Let us begin by observing that if  $r > 0$  then we have the lower bound  $r_i \geq r$ . Otherwise, let us fix any  $r > 0$  and restrict ourselves to the collection of balls in  $U_>^a$  with  $r_i \geq r$ . There is no loss in this so long as the estimates we will prove will be independent of our choice of  $r$ , and thus by letting  $r \rightarrow 0$  we will obtain estimates on all of  $U_>^a$ .

Now let us make the precise statement we will prove in this subsection. Namely, consider any of ball centers  $\{x_i\}_{i \in I^a}$  and any radius  $2^{-4}r_i \leq r_\alpha \leq 2$ , where  $r_\alpha = 2^{-\alpha}$ . Then we will show that

$$\mu^a(B_{r_\alpha}(x_i)) \leq 2^{5n}D(n)r_\alpha^k, \quad (8.25)$$

where  $D(n)$  is the constant from the discrete Reifenberg theorem. Let us observe that once we have proved (8.25) then we have finished the proof of the Covering Lemma, as we will then have the estimate

$$\sum r_i^k = \mu^a(B_2(x_i)) \leq C(n). \quad (8.26)$$

We prove (8.25) inductively on  $\alpha$ . To begin notice that for each  $x_i$  if  $\beta$  is the largest integer such that  $2^{-4}r_i \leq r_\beta$ , then the statement clearly holds, by the definition of the measure  $\mu^a$ . In fact, we can go further than this. For each  $x_i$ , let  $r'_i \in [r_i/10, 10r_i]$  be the largest radius such that for all  $10^{-1}r_i \leq s \leq r'_i$  we have

$$\mu^a(B_s(x_i)) \leq 2\omega_k s^k. \quad (8.27)$$

In particular, we certainly have the much weaker estimate  $\mu^a(B_s(x_i)) \leq 2^{5n}D(n)s^k$ , and hence (8.25) is also satisfied for all  $2^{-4}r_i \leq r_\beta \leq r'_i$ . Notice that we then also have the estimate

$$\omega_k r_i^k \leq \mu^a(B_{r_i/10}(x_i)) \leq \mu^a(B_{r'_i}(x_i)) \leq 2\omega_k (r'_i)^k. \quad (8.28)$$

Now let us focus on proving the inductive step of (8.25). Namely, assume  $\alpha$  is such that for all  $x_i$  with  $2^{-4}r_i \leq r_{\alpha+1} < 2$  we have that (8.25) holds. Then we want to prove that the same estimate holds for  $r_\alpha$ . Let us begin by seeing that a weak version of (8.25) holds. Namely, for any index  $i \in I^a$  and any radius  $r_\alpha \leq s \leq 8r_\alpha$ , by covering  $B_s(x_i)$  by at most  $2^{8n}$  balls  $\{B_{r_{\alpha+1}}(y_j)\}$  of radius  $r_{\alpha+1}$  we have the weak estimate

$$\mu^a(B_s(x_i)) \leq \sum \mu^a(B_{r_{\alpha+1}}(y_j)) \leq D'(n)s^k, \quad (8.29)$$

where of course  $D'(n) \gg 2^{5n}D(n)$ .

To improve on this, let us fix an  $i \in I^a$  and the relative ball center  $x_i \in U_>^a$  with  $2^{-4}r_i \leq r_\alpha$ . Now let  $\{x_j\}_{j \in J} = \{x_i\}_{i \in I} \cap B_{r_\alpha}(x_i)$  be the collection of ball centers in  $B_{r_\alpha}(x_i)$ . Notice first that if  $r'_j > 2r_\alpha$  for any of the ball centers  $\{x_j\}$ , then we can estimate

$$\mu^a(B_{r_\alpha}(x_i)) \leq \mu^a(B_{2r_\alpha}(x_j)) \leq 2\omega_k (2r_\alpha)^k \leq 2^{5n}D(n)r_\alpha^k, \quad (8.30)$$

so that we may fairly assume  $r'_j \leq 2r_\alpha$  for every  $j \in J$ . Now for each ball  $B_{r'_j}(x_j)$  let us define a new ball  $B_{\bar{r}_j}(y_j)$  which is roughly equivalent, but will have some additional useful properties needed to apply the

discrete Reifenberg. Namely, for a given ball  $B_{r'_j}(x_j)$ , let us consider the two options  $r'_j < 10r_j$  or  $r'_j = 10r_j$ . If  $r'_j < 10r_j$ , then we let  $y_j \equiv x_j$  with  $\bar{r}_j \equiv r_j$ . In this case we must have that  $\mu^a(B_{10r_j}(x_j)) > 2\omega_n(r_j)^k$ , and thus we can apply Lemma 8.5 in order to conclude that  $\theta_{\eta\bar{r}_j}(y_j) \geq E - \eta'$ . In the case when  $r'_j = 10r_j$  is maximal, let  $y_j \in B_{r_j}(x_j) \cap S_{\epsilon, r}^k$  with  $y_j \neq x_j$  be a point such that  $\theta_{\eta r_j}(y_j) = E - \eta$ , such a point exists by the definition of  $r_j$ , and let  $\bar{r}_j \equiv 9r_j$ . In either case we then have the estimates

$$\begin{aligned} \theta_{\eta'\bar{r}_j}(y_j) &\geq E - \eta', \\ \omega_k 10^{-k} \bar{r}_j^k &\leq \mu(B_{\bar{r}_j/10}(y_j)) \leq \mu^a(B_{\bar{r}_j}(y_j)) \leq 4\omega_k \bar{r}_j^k, \\ B_{r_j}(x_j) &\subseteq B_{\bar{r}_j}(y_j). \end{aligned} \quad (8.31)$$

Now let us consider the covering  $U_\alpha^a \cap B_{r_\alpha}(x_i) \subseteq \bigcup B_{\bar{r}_j/10}(y_j)$ , and choose from it a Vitali subcovering

$$U_\alpha^a \cap B_{r_\alpha}(x_i) \subseteq \bigcup B_{\bar{r}_j}(y_j), \quad (8.32)$$

such that  $\{B_{\bar{r}_j/10}(y_j)\}$  are disjoint, where we are now being loose on notation and referring to  $\{y_j\}_{j \in \bar{J}}$  as the ball centers from this subcovering. Let us now consider the measure

$$\mu' \equiv \sum_{j \in \bar{J}} \omega_k \left( \frac{\bar{r}_j}{10} \right)^k \delta_{y_j}. \quad (8.33)$$

That is, we have associated to the disjoint collection  $\{B_{\bar{r}_j/10}(y_j)\}$  the natural measure. Our goal is to prove that

$$\mu'(B_{r_\alpha}(x_i)) \leq D(n)r_\alpha^k. \quad (8.34)$$

Let us observe that if we prove (8.34) then we are done. Indeed, using (8.31) we can estimate

$$\mu^a(B_{r_\alpha}(x_i)) \leq \sum \mu^a(B_{\bar{r}_j}(y_j)) \leq 4\omega_k \sum \bar{r}_j^k = 4 \cdot 10^k \mu'(B_{r_\alpha}(x_i)) \leq 2^{5n} D(n) r_\alpha^k, \quad (8.35)$$

which would finish the proof of (8.25) and therefore the Lemma.

Thus let us concentrate on proving (8.34). We will want to apply the discrete Reifenberg in this case to the measure  $\mu'$ . Let us begin by proving a weak version of (8.34). Namely, for any ball center  $y_j$  from our subcovering and radius  $\bar{r}_j < s \leq 4r_\alpha$  let us consider the set  $\{z_k\} = \{y_s\}_{s \in \bar{J}} \cap B_s(y_j)$  of ball centers inside  $B_s(y_j)$ . Since the balls  $\{B_{\bar{r}_k/5}(z_k)\}$  are disjoint we have that  $\bar{r}_k \leq 10s$ . Using this, (8.28), and (8.31) we can estimate

$$\mu'(B_s(y_j)) = \sum_{z_k \in B_s(y_j)} \omega_k 10^{-k} \bar{r}_k^k \leq C(n) \sum_{z_k \in B_s(y_j)} \mu^a(B_{\bar{r}_k/8}(z_k)) \leq C(n) \mu^a(B_{2s}(y_j)) \leq C(n) s^k, \quad (8.36)$$

where of course  $C(n) \gg 2^{5n} D(n)$ .

Now let us finish the proof of (8.34). Thus let us pick a ball center  $y_j \in B_{r_\alpha}(x_i)$  and a radius  $s < 4r_\alpha$ . Note that by the first equation in (8.31) and theorem 2.3 we have that  $B_s(y_j)$  can be made arbitrarily 0-symmetric by choosing  $\eta'$  sufficiently small. If  $\mu'(B_s(y_j)) \leq \epsilon_n s^k$  then  $D_{\mu'}(y_j, s) \equiv 0$  by definition, and if  $s \leq \bar{r}_j/10$  then  $D_{\mu'}(y, s) \equiv 0$ , since the support of  $\mu'$  in  $B_{\bar{r}_j/10}(y_j)$  contains at most one point and thus is precisely contained

in a  $k$ -dimensional subspace. In the case when  $s > \bar{r}_i/10$  and  $\mu(B_s(y)) > \epsilon_n s^k$  then by Theorem 7.1 we have that

$$D_{\mu'}(y_j, s) \leq C(n, K_N, \Lambda, \epsilon) s^{-k} \int_{B_s(y)} W_s(z) d\mu'(z). \quad (8.37)$$

By applying this to all  $r \leq t \leq s$  we can estimate

$$\begin{aligned} s^{-k} \int_{B_s(x)} D_{\mu'}(y, t) d\mu'(y) &\leq C s^{-k} \int_{B_s(x)} t^{-k} \int_{B_t(y)} W_t(z) d\mu'(z) d\mu'(y) \\ &= C s^{-k} t^{-k} \int_{B_s(x)} \mu'(B_t(y)) W_t(y) d\mu'(y) \\ &\leq C s^{-k} \int_{B_s(x)} W_t(y) d\mu'(y), \end{aligned} \quad (8.38)$$

where we have used our estimate on  $\mu'(B_t(y))$  from (8.36) in the last line. Let us now consider the case when  $t = r_\beta \leq s \leq 2r_\alpha$ . Then we can sum to obtain:

$$\begin{aligned} \sum_{r_\beta \leq s} s^{-k} \int_{B_s(x)} D_{\mu'}(y, r_\beta) d\mu'(y) &\leq C \sum_{r'_y \leq r_\beta \leq s} s^{-k} \int_{B_s(x)} W_{r_\beta}(y) d\mu'(y) \\ &= C s^{-k} \int_{B_s(x)} \sum_{\bar{r}_y \leq r_\beta \leq s} W_{r_\beta}(y) d\mu'(y) \\ &\leq C s^{-k} \int_{B_s(x)} |\theta_{4s}(y) - \theta_{\bar{r}_y}(y)| d\mu'(y) \leq C(n, K_N, \Lambda, \epsilon) \eta', \end{aligned} \quad (8.39)$$

where we are using (8.31) in the last line in order to see that  $|\theta_s(y) - \theta_{\bar{r}_y}(y)| \leq \eta'$ . Now let us choose  $\eta' \leq \eta'(n, K_N, \Lambda, \epsilon)$  such that

$$\sum_{r_\beta \leq s} s^{-k} \int_{B_s(x)} D_{\mu'}(y, r_\beta) d\mu'(y) \leq \delta^2, \quad (8.40)$$

where  $\delta$  is chosen from the discrete rectifiable-Reifenberg of Theorem 3.4. Since the estimate (8.40) holds for all  $B_s \subseteq B_{2r_\alpha}(x)$ , we can therefore apply Theorem 3.4 to conclude the estimate

$$\mu'(B_{r_\alpha}(x'_i)) \leq D(n) r_\alpha^k. \quad (8.41)$$

This finishes the proof of (8.31), and hence the proof of Lemma 8.1 for the sets  $U_r$  and  $U_+$ .  $\square$

**8.3. Estimating  $U_0$  in Lemma 8.1.** We now finish the proof of Lemma 8.1. Let us begin by proving the Minkowski estimates on  $U_0$ . Indeed, observe for any  $r > 0$  that  $U_0 \subseteq U_r$ , and thus we have the estimate

$$\text{Vol}(B_r(U_0)) \leq \text{Vol}(B_r(U_r)) \leq \omega_n r^n \cdot N \leq C(n, K_N, \Lambda, \epsilon) r^{n-k}, \quad (8.42)$$

which proves the Minkowski claim. In particular, we have as a consequence the  $k$ -dimensional Hausdorff measure estimate

$$\lambda^k(U_0) \leq C(n, K_N, \Lambda, \epsilon). \quad (8.43)$$

In fact, let us conclude a slightly stronger estimate, since it will be a convenient technical tool in the remainder of the proof. If  $B_s(x)$  is any ball with  $x \in B_1$  and  $s < \frac{1}{2}$ , then by applying the same proof to the rescaled ball  $B_s(x) \rightarrow B_1(0)$ , we can obtain the Hausdorff measure estimate

$$\lambda^k(U_0 \cap B_r(x)) \leq Cr^k. \quad (8.44)$$

To finish the construction we need to see that  $U_0$  is rectifiable. We will in fact apply Theorem 3.3 in order to conclude this. To begin with, let  $\mu \equiv \lambda^k|_{U_0}$  be the  $k$ -dimensional Hausdorff measure, restricted to  $U_0$ . Let  $B_s(y)$  be a ball with  $y \in B_1$  and  $s < \frac{1}{2}$ , now we will argue in a manner similar to Section 8.2. Thus, if  $\mu(B_s(y)) \leq \epsilon_n s^k$  then  $D_{U_0}(y, s) \equiv 0$ , and by theorem 2.3 and Theorem 7.1 we have that

$$D_\mu(y, s) \leq C(n, K_N, \Lambda, \epsilon) s^{-k} \int_{B_s(y)} W_s(z) d\mu(z). \quad (8.45)$$

By applying this to all  $t \leq s$  we have

$$\begin{aligned} s^{-k} \int_{B_s(x)} D_\mu(y, t) d\mu(y) &\leq C s^{-k} \int_{B_s(x)} t^{-k} \int_{B_t(y)} W_t(z) d\mu(z) d\mu(y) \\ &= C s^{-k} t^{-k} \int_{B_s(x)} \mu(B_t(y)) W_t(y) d\mu(y) \\ &\leq C s^{-k} \int_{B_s(x)} W_t(y) d\mu(y), \end{aligned} \quad (8.46)$$

where we have used our estimate (8.44) in the last line. Let us now consider the case when  $t = r_\beta = 2^{-\beta} \leq s$ . Then we can sum to obtain:

$$\begin{aligned} \sum_{r_\beta \leq s} s^{-k} \int_{B_s(x)} D_\mu(y, r_\beta) d\mu(y) &\leq C \sum_{r_\beta \leq s} s^{-k} \int_{B_s(x)} W_{r_\beta}(y) d\mu(y) \\ &= C s^{-k} \int_{B_s(x)} \sum_{r_\beta \leq s} W_{r_\beta}(y) d\mu(y) \\ &\leq C s^{-k} \int_{B_s(x)} |\theta_s(y) - \theta_0(y)| d\mu(y) \leq C(n, K_N, \Lambda, \epsilon) \eta, \end{aligned} \quad (8.47)$$

where we have used two points in the last line. First, we have used our estimate  $\mu(B_s(x)) \leq C s^k$ . Second, we have used that by the definition of  $U_0$ , for each point in the support of  $\mu$  we have that  $|\theta_s(y) - \theta_0(y)| \leq \eta$ . Now let us choose  $\eta \leq \eta(n, K_N, \Lambda, \epsilon)$  such that we have

$$\sum_{r_\beta \leq s} s^{-k} \int_{B_s(x)} D_\mu(y, r_\beta) d\mu(y) \leq \delta^2, \quad (8.48)$$

where  $\delta$  is chosen from Theorem 3.3. Thus, by applying Theorem 3.3 we see that  $U_0$  is rectifiable, which finishes the proof of Lemma 8.1 in the context of  $U_0$ .  $\square$



## 9. PROOF OF MAIN THEOREMS FOR STATIONARY HARMONIC MAPS

In this section we prove the main theorems of the paper concerning stationary harmonic maps. With the tools of Sections 3, 7, and 8 developed, we will at this stage mainly be applying the covering of Lemma 8.1 iteratively to arrive at the estimates. When this is done carefully, we can combine the covering lemma with the cone splitting in order to check that for  $k$ -a.e.  $x \in S_\epsilon^k$  there exists a unique  $k$ -dimensional subspace  $V^k \subseteq T_x M$  such that every tangent map of  $f$  at  $x$  is  $k$ -symmetric with respect to  $V$ .

For the proofs of the Theorems' of this section let us make the following remark. For any  $\delta > 0$  we can cover  $B_1(p)$  by a collection of balls

$$B_1(p) \subseteq \bigcup_1^N B_{K_M^{-1}\delta}(p_i), \quad (9.1)$$

where  $N \leq C(n, K_M)\delta^{-n}$ . Thus if  $\delta = \delta(n, K_M, K_N, \Lambda, \epsilon)$  and we can analyze each such ball, then this gives us estimates on all of  $B_1(p)$ . In particular, by rescaling  $B_{K_M^{-1}\delta}(p_i) \rightarrow B_1(p_i)$ , we see that we can assume in our analysis that  $K_M < \delta$  without any loss of generality. We shall do this throughout this section.

**9.1. Proof of Theorem 1.3.** Let  $f : B_2(p) \subseteq M \rightarrow N$  be a stationary harmonic map satisfying (1.16) with  $\int_{B_2(p)} |\nabla f|^2 \leq \Lambda$ . With  $\epsilon, r > 0$  fixed, let us choose  $\eta(n, K_N, \Lambda, \epsilon) > 0$  and  $\delta(n, K_N, \Lambda, \epsilon) > 0$  as in Lemma 8.1. By the remarks around (9.1) we see that we can assume that  $K_M < \delta$ , which we will do for the remainder of the proof.

Now let us begin by first considering an arbitrary ball  $B_s(x)$  with  $x \in B_1(p)$  and  $r < s \leq 1$ , potentially quite small. We will use Lemma 8.1 in order to build a special covering of  $S_{\epsilon,r}^k \cap B_s(x)$ . Let us define

$$E_{x,s} \equiv \sup_{y \in B_s(x) \cap S_{\epsilon,r}^k} \theta_s(y), \quad (9.2)$$

and thus if we apply Lemma 8.1 with  $\eta(n, K_N, \Lambda, \epsilon)$  fixed to  $B_s(x)$ , then we can build a covering

$$S_{\epsilon,r}^k \cap B_s(x) \subseteq U_r \cup U_+ = \bigcup B_r(x'_i) \cup \bigcup B_{r_i}(x_i), \quad (9.3)$$

with  $r_i > r$ . Let us recall that this covering satisfies the following:

- (a)  $r^{k-n} \text{Vol}(B_r U'_r) + \omega_k \sum r_i^k \leq C(n) s^k$ .
- (b)  $\sup_{y \in B_{r_i}(x_i)} \theta_{r_i}(y) \leq E_{x,s} - \eta$ .

Now that we have built our required covering on an arbitrary ball  $B_s(x)$ , let us use this iteratively to build our final covering of  $S_{\epsilon,r}^k(f)$ . First, let us apply it to  $B_1(p)$  in order to construct a covering

$$S_{\epsilon,r}^k(f) \subseteq U_r^1 \cup U_+^1 = \bigcup B_r(x_i^{r,1}) \cup \bigcup B_{r_i^1}(x_i^1), \quad (9.4)$$

such that

$$r^{k-n} \text{Vol}(B_r(U_r^1)) + \omega_k \sum (r_i^1)^k \leq C(n, K_N, \Lambda, \epsilon), \quad (9.5)$$

and with

$$\sup_{y \in B_{r_i^1}(x_i^1) \cap S_{\epsilon, r}^k} \theta_{r_i^1}(y) \leq \Lambda - \eta. \quad (9.6)$$

Now let us tackle the following claim, which is our main iterative step in the proof:

**Claim:** For each  $\ell$  there exists a constant  $C_\ell(\ell, n, K_M, K_N, \Lambda, \epsilon)$  and a covering

$$S_{\epsilon, r}^k(f) \subseteq U_r^\ell \cup U_+^\ell = \bigcup B_r(x_i^{r, \ell}) \bigcup B_{r_i^\ell}(x_i^\ell), \quad (9.7)$$

with  $r_i^\ell > r$ , such that the following two properties hold:

$$\begin{aligned} r^{k-n} \text{Vol}(B_r(U_r^\ell)) + \omega_k \sum (r_i^\ell)^k &\leq C_\ell(\ell, n, K_N, \Lambda, \epsilon), \\ \sup_{y \in B_{r_i^\ell}(x_i^\ell) \cap S_{\epsilon, r}^k} \theta_{r_i^\ell}(y) &\leq \Lambda - \ell \cdot \eta. \end{aligned} \quad (9.8)$$

To prove the claim let us first observe that we have shown this holds for  $\ell = 1$ . Thus let us assume we have proved the claim for some  $\ell$ , and determine from this how to build the covering for  $\ell + 1$  with some constant  $C_{\ell+1}(\ell + 1, n, K_M, K_N, \Lambda, \epsilon)$ , which we will estimate explicitly.

Thus with our covering determined at stage  $\ell$ , let us apply the covering of (9.3) to each ball  $\{B_{r_i^\ell}(x_i^\ell)\}$  in order to obtain a covering

$$S_{\epsilon, r}^k \cap B_{r_i^\ell}(x_i^\ell) \subseteq U_{i, r} \cup U_{i, +} = \bigcup_j B_r(x_{i, j}^{r, \ell}) \bigcup_j B_{r_{i, j}}(x_{i, j}), \quad (9.9)$$

such that

$$\begin{aligned} r^{k-n} \text{Vol}(B_r(U_{i, r})) + \omega_k \sum_j (r_{i, j})^k &\leq C(n, K_N, \Lambda, \epsilon)(r_i^\ell)^k, \\ \sup_{y \in B_{r_{i, j}}(x_{i, j}) \cap S_{\epsilon, r}^k} \theta_{r_{i, j}}(y) &\leq \Lambda - (\ell + 1)\eta. \end{aligned} \quad (9.10)$$

Let us consider the sets

$$\begin{aligned} U_r^{\ell+1} &\equiv U_r^\ell \bigcup_i U_{i, r}, \\ U_+^{\ell+1} &\equiv \bigcup_{i, j} B_{r_{i, j}}(x_{i, j}). \end{aligned} \quad (9.11)$$

Notice that the second property of (9.8) holds for  $\ell + 1$  by the construction, hence we are left analyzing the volume estimate of the first property. Indeed, for this we combine our inductive hypothesis (9.8) for  $U^\ell$  and

(9.10) in order to estimate

$$\begin{aligned}
r^{k-n}\text{Vol}\left(B_r\left(U_r^{\ell+1}\right)\right) + \omega_k \sum_{i,j} (r_{i,j})^k &\leq r^{k-n}\text{Vol}\left(B_r\left(U_r^\ell\right)\right) + \sum_i \left(r^{k-n}\text{Vol}\left(B_r\left(U_{i,r}\right)\right) + \omega_k \sum_j (r_{i,j})^k\right) \\
&\leq C_\ell + C \sum_i (r_i^\ell)^k \\
&\leq C(n, K_M, K_N, \Lambda, \epsilon) \cdot C_\ell(\ell, n, K_M, K_N, \Lambda, \epsilon) \\
&\equiv C_{\ell+1}.
\end{aligned} \tag{9.12}$$

Hence, we have proved that if the claim holds for some  $\ell$  then the claim holds for  $\ell + 1$ . Since we have already shown the claim holds for  $\ell = 1$ , we have therefore proved the claim for all  $\ell$ .

Now we can finish the proof. Indeed, let us take  $\ell = \lceil \eta^{-1}\Lambda \rceil = \ell(\eta, \Lambda)$ . Then if we apply the Claim to such an  $\ell$ , we must have by the second property of (9.8) that

$$U_+^\ell \equiv \emptyset, \tag{9.13}$$

and therefore we have a covering

$$S_{\epsilon,r}^k \subseteq U_r^\ell = \bigcup_i B_r(x_i). \tag{9.14}$$

But in this case we have by (9.8) that

$$\text{Vol}\left(B_r\left(S_{\epsilon,r}^k(f)\right)\right) \leq \text{Vol}\left(B_r\left(U_r^\ell\right)\right) \leq C(n, K_N, \Lambda, \epsilon)r^{n-k}, \tag{9.15}$$

which proves the Theorem.  $\square$

**9.2. Proof of Theorem 1.4.** There are several pieces to Theorem 1.4. To begin with, the volume estimate follows easily now that Theorem 1.3 has been proved. That is, for each  $r > 0$  we have that

$$S_\epsilon^k(f) \subseteq S_{\epsilon,r}^k(f), \tag{9.16}$$

and therefore we have the volume estimate

$$\text{Vol}\left(B_r\left(S_\epsilon^k(f)\right)\right) \leq \text{Vol}\left(B_r\left(S_{\epsilon,r}^k(f)\right)\right) \leq C(n, K_M, K_N, \Lambda, \epsilon)r^{n-k}. \tag{9.17}$$

In particular, this implies the much weaker Hausdorff measure estimate

$$\lambda^k(S_\epsilon^k(f)) \leq C(n, K_M, K_N, \Lambda, \epsilon), \tag{9.18}$$

which proves the first part of the Theorem.

Let us now focus on the rectifiability of  $S_\epsilon^k$ . We consider the following claim, which is the  $r = 0$  version of the main Claim of Theorem 1.3. We will be applying Lemma 8.1, which requires  $K_M < \delta$ . As in the proof of Theorem 1.3 we can just assume this without any loss, as we can cover  $B_1(p)$  by a controlled number of balls of radius  $K_M^{-1/2}\delta$ , so that after rescaling we can analyze each of these balls with the desired curvature assumption. Thus let us consider the following:

**Claim:** If  $K_M < \delta$ , then for each  $\ell$  there exists a covering  $S_\epsilon^k(f) \subseteq U_0^\ell \cup U_+^\ell = U_0^\ell \cup B_{r_i^\ell}(x_i^\ell)$  such that

- (1)  $\lambda^k(U_0^\ell) + \omega_k \sum (r_i^\ell)^k \leq C_\ell(\ell, n, K_N, \Lambda, \epsilon)$ .
- (2)  $U_0^\ell$  is  $k$ -rectifiable.
- (3)  $\sup_{y \in B_{r_i^\ell}(x_i^\ell) \cap S_\epsilon^k} \theta_{r_i^\ell}(y) \leq \Lambda - \ell \cdot \eta$

The proof of the Claim follows essentially the same steps as those for the main Claim of Theorem 1.3. For base step  $\ell = 0$ , we consider the decomposition  $S_\epsilon^k \subseteq U_0^0 \cup U_+^0$  where  $U_0^0 = \emptyset$  and  $U_+^0 = B_1(p)$ .

Now let us assume we have proved the claim for some  $\ell$ , then we wish to prove the claim for  $\ell + 1$ . Thus, let us consider the set  $U_+^\ell$  from the previous covering step given by

$$U_+^\ell = \bigcup B_{r_i^\ell}(x_i^\ell). \quad (9.19)$$

Now let us apply Lemma 8.1 to each of the balls  $B_{r_i^\ell}(x_i^\ell)$  in order to write

$$S_\epsilon^k \cap B_{r_i^\ell}(x_i^\ell) \subseteq U_{i,0} \cup U_{i,+} = U_{i,0} \cup \bigcup_j B_{r_{i,j}}(x_{i,j}), \quad (9.20)$$

with the following properties:

- (a)  $\lambda^k(U_{i,0}) + \omega_k \sum_j r_{i,j}^k \leq C(n, K_N, \Lambda, \epsilon, p)(r_i^\ell)^k$ ,
- (b)  $\sup_{y \in B_{r_{i,j}}(x_{i,j}) \cap S_\epsilon^k} \theta_{r_{i,j}}(y) \leq \Lambda - (\ell + 1)\eta$ ,
- (c)  $U_{i,0}$  is  $k$ -rectifiable.

Now let us define the sets

$$\begin{aligned} U_0^{\ell+1} &= \bigcup U_{i,0} \cup U_0^\ell, \\ U_+^{\ell+1} &= \bigcup_{i,j} B_{r_{i,j}}(x_{i,j}). \end{aligned} \quad (9.21)$$

Conditions (2) and (3) from the Claim are clearly satisfied. We need only check condition (1). Using (a) and the inductive hypothesis we can estimate that

$$\begin{aligned} \lambda^k(U_0^{\ell+1}) + \omega_k \sum_{i,j} (r_{i,j})^k &\leq \lambda^k(U_0^\ell) + \sum_i \left( \lambda^k(U_{i,0}) + \omega_k \sum_j (r_{i,j})^k \right), \\ &\leq C_\ell + C(n, K_N, \Lambda, \epsilon) \sum_i (r_i^\ell)^k \\ &\leq C(n, K_N, \Lambda, \epsilon) \cdot C_\ell \\ &\equiv C_{\ell+1}. \end{aligned} \quad (9.22)$$

Thus, we have proved the inductive part of the claim, and thus the claim itself.

Let us now finish the proof that  $S_\epsilon^k(f)$  is rectifiable. So let us take  $\ell = \lceil \eta^{-1} \Lambda \rceil = \ell(\eta, \Lambda)$ . Then if we apply the above Claim to  $\ell$ , then by the third property of the Claim we must have that

$$U_+^\ell \equiv \emptyset, \quad (9.23)$$

and therefore we have the covering

$$S_\epsilon^k \subseteq U_0^\ell, \quad (9.24)$$

where  $U_0^\ell$  is  $k$ -rectifiable with the volume estimate  $\lambda^k(U_0^\ell) \leq C$ , which proves that  $S_\epsilon^k$  is itself rectifiable.

Finally, we prove that for  $k$  a.e.  $x \in S_\epsilon^k$  there exists a  $k$ -dimensional subspace  $V_x \subseteq T_x M$  such that *every* tangent map at  $x$  is  $k$ -symmetric with respect to  $V_x$ . To see this we proceed as follows. For each  $\eta > 0$  let us consider the finite decomposition

$$S_\epsilon^k = \bigcup_{\alpha=0}^{\lceil \eta^{-1} \Lambda \rceil} W_{\epsilon, \eta}^{k, \alpha}, \quad (9.25)$$

where by definition we have

$$W_{\epsilon, \eta}^{k, \alpha} \equiv \{x \in S_\epsilon^k : \theta_0(x) \in [\alpha\eta, (\alpha + 1)\eta)\}. \quad (9.26)$$

Note then that each  $W_{\epsilon, \eta}^{k, \alpha}$  is  $k$ -rectifiable, and thus there exists a full measure subset  $\tilde{W}_{\epsilon, \eta}^{k, \alpha} \subseteq W_{\epsilon, \eta}^{k, \alpha}$  such that for each  $x \in \tilde{W}_{\epsilon, \eta}^{k, \alpha}$  the tangent cone of  $W_{\epsilon, \eta}^{k, \alpha}$  exists and is a subspace  $V_x \subseteq T_x M$ .

Now let us consider such an  $x \in \tilde{W}_{\epsilon, \eta}^{k, \alpha}$ , and let  $V_x^k$  be the tangent cone of  $W_{\epsilon, \eta}^{k, \alpha}$  at  $x$ . For all  $r \ll 1$  sufficiently small we of course have  $|\theta_r(x) - \theta_0(x)| < \eta$ . Thus, by the monotonicity and continuity of  $\theta$  we have for all  $r \ll 1$  sufficiently small and all  $y \in W_{\epsilon, \eta}^{k, \alpha} \cap B_r(x)$  that  $|\theta_r(y) - \theta_0(y)| < 2\eta$ . In particular, by Theorem 2.1 we have for each  $y \in W_{\epsilon, \eta}^{k, \alpha} \cap B_r(x)$  that  $B_r(y)$  is  $(0, \delta_\eta)$ -symmetric, with  $\delta_\eta \rightarrow 0$  as  $\eta \rightarrow 0$ . Now let us recall the cone splitting of Theorem 2.3. Since the tangent cone at  $x$  is  $V_x^k$ , for all  $r$  sufficiently small we can find  $k + 1$  points  $x_0, \dots, x_k \in B_r(x) \cap W_{\epsilon, \eta}^{k, \alpha}$  which are  $10^{-1}r$ -independent, see Definition 2.2, and for which  $B_{2r}(x_j)$  are  $(0, \delta_\eta)$ -symmetric. Thus, by the cone splitting of Theorem 2.3 we have that  $B_r(x)$  is  $(k, \delta_\eta)$ -symmetric with respect to  $V_x^k$  for all  $r$  sufficiently small, where  $\delta_\eta \rightarrow 0$  as  $\eta \rightarrow 0$ . In particular, every tangent map at  $x$  is  $(k, \delta_\eta)$ -symmetric with respect to  $V_x$ , where  $\delta_\eta \rightarrow 0$  as  $\eta \rightarrow 0$ .

Now let us consider the sets

$$\tilde{W}_{\epsilon, \eta}^k \equiv \bigcup_{\alpha} \tilde{W}_{\epsilon, \eta}^{k, \alpha}. \quad (9.27)$$

So  $\tilde{W}_{\epsilon, \eta}^k \subseteq S_\epsilon^k$  is a subset of full  $k$ -dimensional measure, and for every point  $x \in \tilde{W}_{\epsilon, \eta}^k$  we have seen that *every* tangent map of is  $(k, \delta_\eta)$ -symmetric with respect to some  $V_x \subseteq T_x M$ , where  $\delta_\eta \rightarrow 0$  as  $\eta \rightarrow 0$ . Finally let us define the set

$$\tilde{S}_\epsilon^k \equiv \bigcap_j \tilde{W}_{\epsilon, j^{-1}}^k. \quad (9.28)$$

This is a countable intersection of full measure sets, and thus  $\tilde{S}_\epsilon^k \subseteq S_\epsilon^k$  is a full measure subset. Further, we have for each  $x \in \tilde{S}_\epsilon^k$  that *every* tangent map must be  $(k, \delta)$ -symmetric with respect to some  $V_x$ , for all  $\delta > 0$ . In particular, every tangent map at  $x$  must be  $(k, 0) = k$ -symmetric with respect to some  $V_x$ . This finishes the proof of the Theorem.  $\square$

9.3. **Proof of Theorem 1.5.** Let us begin by observing the equality

$$S^k(f) = \bigcup_{\epsilon > 0} S_\epsilon^k(f) = \bigcup_{\beta \in \mathbb{N}} S_{2^{-\beta}}^k(f). \quad (9.29)$$

Indeed, if  $x \in S_\epsilon^k(f)$ , then no tangent map at  $x$  can be  $(k+1, \epsilon/2)$ -symmetric, and in particular  $k+1$ -symmetric, and thus  $x \in S^k(f)$ . This shows that  $S_\epsilon^k(f) \subseteq S^k(f)$ . On the other hand, if  $x \in S^k(f)$  then we claim there is some  $\epsilon > 0$  for which  $x \in S_\epsilon^k(f)$ . Indeed, if this is not the case, then there exists  $\epsilon_i \rightarrow 0$  and  $r_i > 0$  such that  $B_{r_i}(x)$  is  $(k+1, \epsilon_i)$ -symmetric. If  $r_i \rightarrow 0$  then we can pass to a subsequence to find a tangent map which is  $k+1$ -symmetric, which is a contradiction. On the other hand, if  $r_i > r > 0$  then we see that  $B_r(x)$  is itself  $k+1$ -symmetric, and in particular *every* tangent map at  $x$  is  $k+1$ -symmetric. In either case we obtain a contradiction, and thus  $x \in S_\epsilon^k(f)$  for some  $\epsilon > 0$ . Therefore we have proved (9.29).

As a consequence,  $S^k(f)$  is a countable union of  $k$ -rectifiable sets, and therefore is itself  $k$ -rectifiable. On the other hand, Theorem 1.4 tells us that for each  $\beta \in \mathbb{N}$  there exists a set  $\tilde{S}_{2^{-\beta}}^k(f) \subseteq S_{2^{-\beta}}^k(f)$  of full measure such that

$$\tilde{S}_{2^{-\beta}}^k \subseteq \{x : \exists V^k \subseteq T_x M \text{ s.t. every tangent map at } x \text{ is } k\text{-symmetric wrt } V\}. \quad (9.30)$$

Hence, let us define

$$\tilde{S}^k(f) \equiv \bigcup_{\beta \in \mathbb{N}} \tilde{S}_{2^{-\beta}}^k(f). \quad (9.31)$$

Then we still have that  $\tilde{S}^k(f)$  has  $k$ -full measure in  $S^k(f)$ , and if  $x \in \tilde{S}^k(f)$  then for some  $\beta$  we have that  $x \in \tilde{S}_{2^{-\beta}}^k$ , which proves that there exists a subspace  $V \subseteq T_x M$  such that every tangent map at  $x$  is  $k$ -symmetric with respect to  $V$ . We have finished the proof of the theorem.  $\square$

## 10. PROOF OF MAIN THEOREMS FOR MINIMIZING HARMONIC MAPS

In this section we prove the main theorems of the paper concerning minimizing harmonic maps. That is, we finish the proofs of Theorem 1.6 and Theorem 1.8. In fact, the proofs of these two results are almost identical, though the first relies on Theorem 1.4 and the latter on Theorem 1.3. However, for completeness sake we will include the details of both.

10.1. **Proof of Theorem 1.6.** We wish to understand better the size of the singular set  $S(f)$  of a minimizing harmonic map. Let us recall that the  $\epsilon$ -regularity of theorem 2.8 tells us that if  $f$  is minimizing, then there exists  $\epsilon(n, K_N, \Lambda) > 0$  with the property that if  $x \in B_1(p)$  and  $0 < r < r(n, K_M, K_N, \Lambda)$  is such that  $B_{2r}(x)$  is  $(n-2, \epsilon)$ -symmetric, then  $r_f(x) \geq r$ . In particular,  $x$  is a smooth point, and we have for  $\epsilon(n, K_M, K_N, \Lambda) > 0$  that

$$\text{Sing}(f) \cap B_1(p) \subseteq S_\epsilon^{n-3}(f). \quad (10.1)$$

Thus by Theorem 1.4 there exists  $C(n, K_M, K_N, \Lambda) > 0$  such that for each  $0 < r < 1$  we have

$$\text{Vol}(B_r(\text{Sing}(f)) \cap B_1(p)) \leq \text{Vol}(B_r(S_\epsilon^{n-3}(f))) \leq Cr^3. \quad (10.2)$$

This of course immediately implies, though of course is much stronger than, the Hausdorff measure estimate

$$\lambda^{n-3}(\text{Sing}(f) \cap B_1(p)) \leq C, \quad (10.3)$$

which finishes the proof of the Theorem.  $\square$

**10.2. Proof of Theorem 1.8.** We begin again by considering the  $\epsilon$ -regularity of theorem 2.8. This tells us that if  $f$  is minimizing, then there exists  $\epsilon(n, K_N, \Lambda) > 0$  with the property that if  $x \in B_1(p)$  and  $0 < r < r(n, K_M, K_N, \Lambda)$  is such that  $B_{2r}(x)$  is  $(n-2, \epsilon)$ -symmetric, then  $r_f(x) \geq r$ . In particular, we have for such  $\epsilon, r$  that

$$\{x \in B_1(p) : r_f(x) < r\} \subseteq S_{\epsilon, r}^{n-3}(f). \quad (10.4)$$

Thus by Theorem 1.4 there exists  $C(n, K_M, K_N, \Lambda) > 0$  such that for each  $0 < r < 1$  we have

$$\text{Vol}(B_r \{x \in B_1(p) : r_f(x) < r\}) \leq \text{Vol}(B_r(S_{\epsilon, r}^{n-3}(f))) \leq Cr^3, \quad (10.5)$$

which proves the second estimate of (1.23). To prove the first we observe that  $|\nabla f|(x) \leq r_f(x)^{-1}$ , and to prove (1.24) we use the remark after Definition 1.7 to conclude that  $|\nabla^2 f|(x) \leq C(n, K_M, K_N)r_f(x)^{-2}$ . This concludes the proof of the Theorem.  $\square$

## 11. SHARPNESS OF THE RESULTS

In this section we present a few examples which motivate the sharpness of our results.

**11.1. Sharpness of  $L^p$  Estimates for Minimizers.** This example is completely standard, we wish to simply point out some of its properties. Namely, consider the mapping  $f : B_2(0^3) \rightarrow S^2$  given by the projection

$$f(x) = \frac{x}{|x|}. \quad (11.1)$$

This is a minimizing harmonic map (see [Lin87, CG89]). It is easy to compute that

$$|\nabla f|(x) \approx r_f^{-1}(x) \approx \frac{1}{|x|}. \quad (11.2)$$

We therefore get that  $|\nabla f|, r_f^{-1}(x)$  have uniform estimates in  $L_{weak}^3$ , however neither belong to  $L^3$ . In particular, this shows that the estimates of Theorem 1.8 are sharp.

**11.2. Rectifiable-Reifenberg Example I.** Let us begin with an easy example, which shows that the rectifiable conclusions of Theorem 3.3 is sharp. That is, one cannot hope for better structural results under the hypothesis. Indeed, consider any  $k$ -dimensional subspace  $V^k \subseteq \mathbb{R}^n$ , and let  $S \subseteq V^k \cap B_2(0^n)$  be an arbitrary measurable subset. Then clearly  $D(x, r) \equiv 0$  for each  $x$  and  $r > 0$ , and thus the hypotheses of Theorem 3.3 are satisfied, however  $S$  clearly need not be better than rectifiable. In the next example we shall see that  $S$  need not even come from a single rectifiable chart, as it does in this example.

**11.3. Rectifiable-Reifenberg Example II.** With respect to the conclusions of Theorem 3.3 there are two natural questions regarding how sharp they are. First, is it possible to obtain more structure from the set  $S$  than rectifiable? In particular, in Theorem 3.2 there are topological conclusions about the set, is it possible to make such conclusions in the context of Theorem 3.3? In the last example we saw this is not the case. Then a second question is to ask whether we can at least find a single rectifiable chart which covers the whole set  $S$ . This example taken from [DT12, counterexample 12.4] shows that the answer to this question is negative as well.

To build our examples let us first consider a unit circle  $S^1 \subseteq \mathbb{R}^3$ . Let  $M^2 \supset S^1$  be a smooth Möbius strip around this circle, and let  $S_\epsilon \subseteq M^2 \cap B_\epsilon(S^1) \equiv M_\epsilon^2$  be an arbitrary  $\lambda^2$ -measurable subset of the Möbius strip, contained in a small neighborhood of the  $S^1$ . In particular,  $Area(S_\epsilon) \leq C\epsilon \rightarrow 0$  as  $\epsilon \rightarrow 0$ . It is not hard, though potentially a little tedious, to check that assumptions of Theorem 3.3 hold for  $\delta \rightarrow 0$  as  $\epsilon \rightarrow 0$ .

However, we have learned two points from these example. First, since  $S_\epsilon$  was an arbitrary measurable subset of a two dimensional manifold, we have that it is 2-rectifiable, however that is the most which may be said of  $S_\epsilon$ . That is, structurally speaking we cannot hope to say better than 2-rectifiable about the set  $S_\epsilon$ . More than that, since  $S_\epsilon$  is a subset of the Möbius strip, we see that even though  $S_\epsilon$  is rectifiable, we cannot even cover  $S_\epsilon$  by a single chart from  $B_1(0^2)$ , as there are topological obstructions, see [DT12] for more on this.

#### ACKNOWLEDGMENTS

We would like to thank prof. **Camillo de Lellis** for his very precise comments on this paper and its earlier versions.

#### REFERENCES

- [Alm68] F. J. ALMGREN, JR., Existence and regularity almost everywhere of solutions to elliptic variational problems among surfaces of varying topological type and singularity structure, *Ann. of Math. (2)* **87** (1968), 321–391. MR 0225243. Zbl 0162.24703.
- [Alm00] F. J. ALMGREN, JR., *Almgren’s big regularity paper*, *World Scientific Monograph Series in Mathematics* **1**, World Scientific Publishing Co., Inc., River Edge, NJ, 2000,  $Q$ -valued functions minimizing Dirichlet’s integral and the regularity of area-minimizing rectifiable currents up to codimension 2, With a preface by Jean E. Taylor and Vladimir Scheffer. MR 1777737. Zbl 0985.49001.
- [AFP00] L. AMBROSIO, N. FUSCO, and D. PALLARA, *Functions of bounded variation and free discontinuity problems*, *Oxford Mathematical Monographs*, The Clarendon Press, Oxford University Press, New York, 2000. MR 1857292. Zbl 0957.49001.
- [AT15] J. AZZAM and X. TOLSA, Characterization of  $n$ -rectifiability in terms of Jones’ square function: Part II, *Geom. Funct. Anal.* **25** (2015), 1371–1412. MR 3426057. Zbl 06521333. <http://dx.doi.org/10.1007/s00039-015-0334-7>.
- [BMT13] Z. M. BALOGH, R. MONTI, and J. T. TYSON, Frequency of Sobolev and quasiconformal dimension distortion, *J. Math. Pures Appl. (9)* **99** (2013), 125–149. MR 3007840. Zbl 1266.28003. <http://dx.doi.org/10.1016/j.matpur.2012.06.005>.
- [Bet93] F. BETHUEL, On the singular set of stationary harmonic maps, *Manuscripta Math.* **78** (1993), 417–443. MR 1208652. Zbl 0792.53039. <http://dx.doi.org/10.1007/BF02599324>.



- [BL15] C. BREINER and T. LAMM, Quantitative stratification and higher regularity for biharmonic maps, *Manuscripta Math.* **148** (2015), 379–398. MR 3414482. Zbl 1327.53079. <http://dx.doi.org/10.1007/s00229-015-0750-x>. Available at <http://arxiv.org/abs/1410.5640>.
- [CHN13a] J. CHEEGER, R. HASLHOFER, and A. NABER, Quantitative stratification and the regularity of harmonic map flow, *accepted to Calc. of Var. and PDE* (2013). Available at <http://arxiv.org/abs/1308.2514>.
- [CHN13b] ———, Quantitative stratification and the regularity of mean curvature flow, *Geom. Funct. Anal.* **23** (2013), 828–847. MR 3061773. Zbl 1277.53064. <http://dx.doi.org/10.1007/s00039-013-0224-9>. Available at <http://arxiv.org/abs/1308.2514>.
- [CN13a] J. CHEEGER and A. NABER, Lower bounds on Ricci curvature and quantitative behavior of singular sets, *Invent. Math.* **191** (2013), 321–339. MR 3010378. Zbl 1268.53053. <http://dx.doi.org/10.1007/s00222-012-0394-3>. Available at <http://arxiv.org/abs/1103.1819>.
- [CN13b] ———, Quantitative stratification and the regularity of harmonic maps and minimal currents, *Comm. Pure Appl. Math.* **66** (2013), 965–990. MR 3043387. Zbl 1269.53063. <http://dx.doi.org/10.1002/cpa.21446>. Available at <http://arxiv.org/abs/1107.3097>.
- [CNV15] J. CHEEGER, A. NABER, and D. VALTORTA, Critical sets of elliptic equations, *Comm. Pure Appl. Math.* **68** (2015), 173–209. MR 3298662. Zbl 06399723. <http://dx.doi.org/10.1002/cpa.21518>. Available at <http://arxiv.org/abs/1207.4236>.
- [CG89] J.-M. CORON and R. GULLIVER, Minimizing  $p$ -harmonic maps into spheres, *J. Reine Angew. Math.* **401** (1989), 82–100. MR 1018054. Zbl 0677.58021. <http://dx.doi.org/10.1515/crll.1989.401.82>.
- [DS93] G. DAVID and S. SEMMES, *Analysis of and on uniformly rectifiable sets*, *Mathematical Surveys and Monographs* **38**, American Mathematical Society, Providence, RI, 1993. MR 1251061. Zbl 0832.42008. <http://dx.doi.org/10.1090/surv/038>.
- [DT12] G. DAVID and T. TORO, Reifenberg parameterizations for sets with holes, *Mem. Amer. Math. Soc.* **215** (2012), vi+102. MR 2907827. Zbl 1236.28002. <http://dx.doi.org/10.1090/S0065-9266-2011-00629-5>.
- [EG92] L. C. EVANS and R. F. GARIEPY, *Measure theory and fine properties of functions*, *Studies in Advanced Mathematics*, CRC Press, Boca Raton, FL, 1992. MR 1158660. Zbl 0804.28001.
- [Fed69] H. FEDERER, *Geometric measure theory*, *Die Grundlehren der mathematischen Wissenschaften, Band 153*, Springer-Verlag New York Inc., New York, 1969. MR 0257325. Zbl 0176.00801.
- [FMS15] M. FOCARDI, A. MARCHESI, and E. SPADARO, Improved estimate of the singular set of Dir-minimizing  $Q$ -valued functions via an abstract regularity result, *J. Funct. Anal.* **268** (2015), 3290–3325. MR 3336726. Zbl 06475693. <http://dx.doi.org/10.1016/j.jfa.2015.02.011>.
- [HL90] R. HARDT and F.-H. LIN, The singular set of an energy minimizing map from  $B^4$  to  $S^2$ , *Manuscripta Math.* **69** (1990), 275–289. MR 1078359. Zbl 0713.58006. <http://dx.doi.org/10.1007/BF02567926>.
- [Lin87] F.-H. LIN, A remark on the map  $x/|x|$ , *C. R. Acad. Sci. Paris Sér. I Math.* **305** (1987), 529–531. MR 916327. Zbl 0652.58022.
- [Lin99] ———, Gradient estimates and blow-up analysis for stationary harmonic maps, *Ann. of Math. (2)* **149** (1999), 785–829. MR 1709303. Zbl 0949.58017. <http://dx.doi.org/10.2307/121073>. Available at <http://arxiv.org/abs/math/9905214>.
- [LW06] F. H. LIN and C. Y. WANG, Stable stationary harmonic maps to spheres, *Acta Math. Sin. (Engl. Ser.)* **22** (2006), 319–330. MR 2214353. Zbl 1121.58017. <http://dx.doi.org/10.1007/s10114-005-0673-7>.
- [Mat95] P. MATTILA, *Geometry of sets and measures in Euclidean spaces*, *Cambridge Studies in Advanced Mathematics* **44**, Cambridge University Press, Cambridge, 1995, Fractals and rectifiability. MR 1333890. Zbl 0819.28004. <http://dx.doi.org/10.1017/CB09780511623813>.
- [Mor66] C. B. MORREY, JR., *Multiple integrals in the calculus of variations*, *Die Grundlehren der mathematischen Wissenschaften, Band 130*, Springer-Verlag New York, Inc., New York, 1966. MR 0202511. Zbl 0142.38701.
- [NV] A. NABER and D. VALTORTA, Volume estimates on the critical sets of solutions to elliptic pdes, *Preprint*. Available at <http://arxiv.org/abs/1403.4176>.

- [Rei60] E. R. REIFENBERG, Solution of the Plateau Problem for  $m$ -dimensional surfaces of varying topological type, *Acta Math.* **104** (1960), 1–92. MR 0114145. Zbl 0099.08503. Available at <http://link.springer.com/article/10.1007%2FBBF02547186>.
- [SU82] R. SCHOEN and K. UHLENBECK, A regularity theory for harmonic maps, *J. Differential Geom.* **17** (1982), 307–335. MR 664498. Zbl 0521.58021. Available at <http://projecteuclid.org/getRecord?id=euclid.jdg/1214436923>.
- [Sim] L. SIMON, Reifenberg’s topological disc theorem. Available at <http://www.math.uni-tuebingen.de/ab/analysis/pub/leon/reifenberg/reifenberg.dvi>.
- [Sim83] ———, *Lectures on geometric measure theory, Proceedings of the Centre for Mathematical Analysis, Australian National University* **3**, Australian National University Centre for Mathematical Analysis, Canberra, 1983. MR 756417. Zbl 0546.49019.
- [Sim96] ———, *Theorems on regularity and singularity of energy minimizing maps, Lectures in Mathematics ETH Zürich*, Birkhäuser Verlag, Basel, 1996, Based on lecture notes by Norbert Hungerbühler. MR 1399562. Zbl 0864.58015. <http://dx.doi.org/10.1007/978-3-0348-9193-6>.
- [Tol15] X. TOLSA, Characterization of  $n$ -rectifiability in terms of Jones’ square function: part I, *Calc. Var. Partial Differential Equations* **54** (2015), 3643–3665. MR 3426090. Zbl 06544048. <http://dx.doi.org/10.1007/s00526-015-0917-z>.
- [Tor95] T. TORO, Geometric conditions and existence of bi-Lipschitz parameterizations, *Duke Math. J.* **77** (1995), 193–227. MR 1317632. Zbl 0847.42011. <http://dx.doi.org/10.1215/S0012-7094-95-07708-4>.
- [Zie89] W. P. ZIEMER, *Weakly differentiable functions, Graduate Texts in Mathematics* **120**, Springer-Verlag, New York, 1989, Sobolev spaces and functions of bounded variation. MR 1014685. Zbl 0692.46022. Available at <http://www.ams.org/mathscinet-getitem?mr=1014685>.