CHAPTER 2

SOME FURTHER PRELIMINARIES FROM ANALYSIS

Here we develop the necessary further analytical background material needed for later developments. In particular we prove some basic results about Lipschitz and BV functions, and we also present the basic facts concerning C^k submanifolds of Euclidean space. There is also a brief treatment of the area and co-area formula and a discussion of first and second variation formulae for C^2 submanifolds of Euclidean space. These latter topics will be discussed in a much more general context later.

§5. LIPSCHITZ FUNCTIONS

Recall that a function $f : X \to \mathbb{R}$ is said to be Lipschitz if there is $L < \infty$ such that (if d is the metric on X)

$$|f(x) - f(y)| \leq L\, d(x,y) \qquad \forall\, x,y \in X .$$

$\mathrm{Lip}\, f$ denotes the least such constant L.

First we have the following trivial extension theorem.

5.1 THEOREM *If* $A \subset X$ *and* $f : A \to \mathbb{R}$ *is Lipschitz, then* $\exists\, \bar{f} : X \to \mathbb{R}$ *with* $\mathrm{Lip}\, \bar{f} = \mathrm{Lip}\, f$, *and* $f = \bar{f}|A$.

Proof Simply define

$$\bar{f}(x) = \inf_{y\in A}(f(y) + L\, d(x,y)) , \quad L = \mathrm{Lip}\, f .$$

Since $f(y) + L\, d(x,y) \geq f(z) - L\, d(x,z)$ $\forall\, x \in X$, $y,z \in A$, we see that $\bar{f}$

is real-valued and $\bar{f}(x) = f(x)$ for $x \in A$. Furthermore

$$\bar{f}(x_1) - \bar{f}(x_2) = \sup_{y_2 \in A} \inf_{y_1 \in A} (f(y_1) + L\, d(x_1,y_1) - f(y_2) - L\, d(x_2,y_2))$$

$$\leq \sup_{y_2 \in A} (L\, d(x_1,y_2) - L\, d(x_2,y_2))$$

$$\leq L\, d(x_1,x_2) \qquad \forall\, x_1,x_2 \in X\, .$$

Next we need the theorem of Rademacher concerning differentiability of Lipschitz functions on $\mathbb{R}^n$. (The proof given here is due to C.B. Morrey.)

5.2 THEOREM *If* f *is Lipschitz on* $\mathbb{R}^n$ *, then* f *is differentiable* $\mathcal{L}^n$*-almost everywhere; that is,* $\operatorname{grad} f(x) = (D_1 f(x), \ldots, D_n f(x))$ *exists and*

$$\lim_{y \to x} \frac{f(y)-f(x)-\operatorname{grad} f(x)\cdot(y-x)}{|y-x|} = 0$$

for $\mathcal{L}^n$ - a.e. $x \in \mathbb{R}^n$.

Proof Let $v \in S^{n-1}$, and whenever it exists let $D_v f(x)$ denote the directional derivative $\frac{d}{dt} f(x+tv)\Big|_{t=0}$. Since $D_v f(x)$ exists precisely when the Borel-measurable functions $\limsup_{t \to 0} \frac{f(x+tv)-f(x)}{t}$ and $\liminf_{t \to 0} \frac{f(x+tv)-f(x)}{t}$ coincide, the set A_v on which $D_v f$ fails to exist is $\mathcal{L}^n$-measurable. However $\phi(t) = f(x+tv)$ is an absolutely continuous function of $t \in \mathbb{R}$ for any fixed x and v , and hence is differentiable for almost all t . Thus A_v intersects every line L which is parallel to v in a set of $\mathcal{H}^1$ measure zero. Thus for each $v \in S^{n-1}$

(1) $$D_v f(x) \text{ exists } \mathcal{L}^n\text{- a.e. } x \in \mathbb{R}^n\, .$$

Now take any $C_0^\infty(\mathbb{R}^n)$ function ζ and note that for any $h > 0$

$$\int_{\mathbb{R}^n} \frac{f(x+hv)-f(x)}{h} \zeta(x)\,d\mathcal{L}^n(x) = -\int_{\mathbb{R}^n} \frac{\zeta(x)-\zeta(x-hv)}{h} f(x)\,d\mathcal{L}^n(x)$$

(by the change of variable $z = x+hv$ in the first part of the integral on the left). Using the dominated convergence theorem and (1) we then get

$$\begin{aligned}\int D_v f\,\zeta &= -\int f\,D_v\,\zeta = -\int f\,v\cdot\text{grad}\,\zeta \\ &= -\sum_{j=1}^{n} v^j \int f\,D_j\,\zeta = +\sum_{j=1}^{n} v^j \int \zeta\,D_j\,f \\ &= \int \zeta\,v\cdot\text{grad}\,f\,,\end{aligned}$$

where all integrals are with respect to Lebesgue measure on $\mathbb{R}^n$, and where we have used Fubini's theorem and the absolute continuity of f on lines to justify the integration by parts. Since ζ is arbitrary we thus have

$$(2) \qquad D_v f(x) = v\cdot\text{grad}\,f(x)\,,\ \mathcal{L}^n\text{- a.e. } x \in \mathbb{R}^n\,.$$

Now let $v_1, v_2, \ldots$ be a countable dense subset of S^{n-1}, and let $A_k = \{x : \text{grad}\,f(x)\,,\ D_{v_k} f(x)$ exist and $D_{v_k} f(x) = v_k\cdot\text{grad}\,f(x)\}$. Then with $A = \bigcap_{k=1}^{\infty} A_k$ we have by (2) that

$$(3) \quad \mathcal{L}^n(\mathbb{R}^n \sim A) = 0\,,\ D_{v_k} f(x) = v_k\cdot\text{grad}\,f(x) \qquad \forall\, x \in A\,,\ k = 1,2,\ldots\,.$$

Using this, we are now going to prove that f is differentiable at each point x of A. To see this, for any $x \in A$, $v \in S^{n-1}$ and $h > 0$ define

$$Q(x,v,h) = \frac{f(x+hv)-f(x)}{h} - v\cdot\text{grad}\,f(x)\,.$$

Evidently for any $x \in A$, $v,v' \in S^{n-1}$, $h > 0$,

$$(4) \qquad |Q(x,v,h) - Q(x,v',h)| \le (n+1)L\,|v-v'|\,,\quad L = \text{Lip}\,f\,.$$

Now let $\varepsilon > 0$ be given and select P large enough so that

(5) $\quad v \in S^{n-1} \Rightarrow |v-v_k| < \frac{\varepsilon}{2(n+1)L}$ *for some* $k \in \{1,\ldots,P\}$.

Since $\lim_{h\downarrow 0} Q(x,v_\ell,h) = 0$, $\forall\, \ell = 1,2,\ldots, x\in A$, (by (2)), we see that for a given $x_0 \in A$ we can choose $\delta > 0$ so that

(6) $\quad |Q(x_0,v_k,h)| < \varepsilon/2$ whenever $0<h<\delta$ and $k \in \{1,\ldots,P\}$.

Since $|Q(x_0,v,h)| \le |Q(x_0,v_k,h)| + |Q(x_0,v,h) - Q(x_0,v_k,h)|$ for each $k \in \{1,\ldots,P\}$, we then have (by (4), (5), (6)) that

$$|Q(x_0,v,h)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

whenever $v \in S^{n-1}$ and $0 < h < \delta$. Thus the theorem is proved.

Finally we shall need the following consequence of the Whitney Extension Theorem.

5.3 THEOREM *Suppose* $f : \mathbb{R}^n \to \mathbb{R}$ *is Lipschitz. Then for each* $\varepsilon > 0$ *there is a* $C^1(\mathbb{R})$ *function* g *with*

$$\mathcal{L}^n(\{x : f(x) \neq g(x)\} \cup \{x : \text{grad } f(x) \neq \text{grad } g(x)\}) < \varepsilon .$$

Proof First recall Whitney's extension theorem for C^1 functions:

If $A \subset \mathbb{R}^n$ is closed and if $h : A \to \mathbb{R}$ and $\nu : A \to \mathbb{R}^n$ are continuous, and if

(*) $$\lim_{\substack{x\to x_0,\, y\to x_0\\ x,y\in A,\, x\neq y}} R(x,y) = 0 \quad \forall\, x_0 \in A ,$$

where

(**) $$R(x,y) = \frac{h(y)-h(x)-\nu(x)\cdot(y-x)}{|x-y|} ,$$

then there is a C^1 function $g : \mathbb{R}^n \to \mathbb{R}$ such that $g = h$ and $\text{grad } g = \nu$ on A . (For the proof see for example [SE] or [FH1]; for the case $n=1$, see Remark 5.4(3) below.)

Now by Rademacher's Theorem f is differentiable on a set $B \subset \mathbb{R}^n$ with $L^n(\mathbb{R}^n \sim B) = 0$. By Lusin's theorem (which applies to sets of infinite measure for L^n) there is a closed set $C \subset B$ such that $\operatorname{grad} f|C$ is continuous and $L^n(\mathbb{R}^n \sim C) < \varepsilon/2$. On C we define $h(x) = f(x)$, $\nu(x) = \operatorname{grad} f(x)$ and $R(x,y)$ for $x,y \in C$ is as defined in (**). Evidently (since $C \subset B$) we have $\lim_{\substack{y \to x \\ y \in C}} R(x,y) = 0 \quad \forall\, x \in C$, but *not* necessarily (*). We therefore proceed as follows. For each $k = 1,2,\ldots$ let

$$\eta_k(x) = \sup\{|R(x,y)| : y \in C \cap (B_{\frac{1}{k}}(x) \sim \{x\})\} .$$

Then $\eta_k \downarrow 0$ pointwise in C, and hence by *Egoroff's Theorem* there is a closed set $A \subset C$ such that $L^n(C \sim A) < \varepsilon/2$ and η_k converges uniformly to zero on A. One now readily checks that (*) holds. Hence we can apply the Whitney Theorem.

5.4 REMARKS

(1) The reader will see that without any significant change the above proof establishes the following: *If* $U \subset \mathbb{R}^n$ *is open and if* $f : U \to \mathbb{R}$ *is differentiable* L^n-a.e. *in* U, *then for each* $\varepsilon > 0$ *there is a closed set* $A \subset U$ *and a* C^1 *function* $g : \mathbb{R}^n \to \mathbb{R}$ *such that* $L^n(U \sim A) < \varepsilon$ *and* $f(x) = g(x)$, $\operatorname{grad} f(x) = \operatorname{grad} g(x)$ *for each* $x \in A$.

(2) The hypothesis (*) above cannot be weakened to the requirement that $\lim_{\substack{y \to x \\ y \in A}} R(x,y) = 0 \quad \forall\, x \in A$. For instance we have the example (for $n = 1$) when $A = \{0\} \cup \{\frac{1}{k} : k = 1,2,\ldots\}$ and $h(0) = 0$, $h(\frac{1}{k}) = (-1)^k/k^{3/2}$, $\nu \equiv 0$. Evidently in this case we have $\lim_{\substack{y \to x \\ y \in A}} R(x,y) = 0 \quad \forall\, x \in A$, but there is no C^1 extension because $\dfrac{|h(\frac{1}{k}) - h(\frac{1}{k+1})|}{(\frac{1}{k} - \frac{1}{k+1})} \to \infty$ as $k \to \infty$.

(3) In the case $n = 1$, the Whitney Extension Theorem used above has a simple proof. Namely in this case define

$$R(x,y) = \frac{h(y) - h(x)}{y - x} - \nu(x)$$

and note that the hypothesis (*) guarantees that for each compact subset C of A we have a function ε_C with $\varepsilon_C(t) \downarrow 0$ as $t \downarrow 0$, and

(i) $$|R(x,y)| \le \varepsilon_C(|x-y|) \quad \forall\, x,y \in C .$$

Notice in particular this implies

(ii) $$|\nu(x) - \nu(y)| \le 2\, \varepsilon_C(|x-y|) \quad \forall\, x,y \in C .$$

Also $\mathbb{R} \sim A$ is a countable disjoint union of open intervals $I_1, I_2, \ldots$. If $I_j = (a,b)$, we then select $g_j \in C^1([a,b])$ as follows:

(iii) $g_j(a) = h(a)$, $g_j(b) = h(b)$, $g_j'(a) = \nu(a)$, $g_j'(b) = \nu(b)$

and

(iv) $$\sup_{x\in I_j} |g_j'(x) - \nu(a)| \le 3\, \varepsilon_C(b-a) , \quad C = [a-1,b+1] \cap A .$$

This is possible by (i), (ii), with $(x,y) = (a,b)$. One now defines $g(x) = g_j(x)$ $\forall\, x \in I_j$, $j = 1,2,\ldots$, and $g(x) = h(x)$ $\forall\, x \in A$. It is then easy to check $g \in C^1(\mathbb{R})$ and $g' = \nu$ on A by virtue of (i) - (iv).

§6. BV FUNCTIONS

In this section we gather together the basic facts about locally BV functions which will be needed later.

First recall that if U is open in $\mathbb{R}^n$ and if $u \in L^1_{loc}(U)$, then u is said to be in $BV_{loc}(U)$ if for each $W \subset\subset U$ there is a constant $c(W) < \infty$ such that

$$\int_W u \operatorname{div} g \, d\mathcal{L}^n \le c(W) \sup|g|$$

for all vector functions $g = (g^1, \dots, g^n)$, $g^j \in C_c^\infty(W)$. Notice that this means that the functional $\int_U u \operatorname{div} g$ extends uniquely to give a (real-valued) linear functional on $K(U, \mathbb{R}^n) \equiv \{$continuous $g = (g^1, \dots, g^n) : U \to \mathbb{R}^n$, support $|g|$ compact$\}$ which is bounded on $K_W(U, \mathbb{R}^n) \equiv \{g \in K(U, \mathbb{R}^n) : \operatorname{spt}|g| \subset W\}$ for every $W \subset\subset U$. Then, by the Riesz representation theorem 4.1, there is a Radon measure μ on U and a μ-measurable function $\nu = (\nu^1, \dots, \nu^n)$, $|\nu| = 1$ a.e., such that

$$\text{6.1} \qquad \int_U u \operatorname{div} g \, d\mathcal{L}^n = \int_U g \cdot \nu \, d\mu .$$

Thus, in the language of distribution theory, the generalized derivatives $D_j u$ of u are represented by the *signed measures* $\nu_j \, d\mu$, $j = 1, \dots, n$. For this reason we often denote the total variation measure μ (see 4.2) by $|Du|$. (In fact if $u \in W^{1,1}_{loc}(U)$ we evidently do have $d\mu = |Du| \, d\mathcal{L}^n$ and

$$\nu_j = \begin{cases} \dfrac{D_j u}{|Du|} & \text{if } |Du| \neq 0 \\ 0 & \text{if } |Du| = 0 \, . \end{cases})$$

Thus for $u \in BV_{loc}(U)$, $|Du|$ will denote the Radon measure on U which is uniquely characterized by

$$|Du|(W) = \sup_{\substack{|g| \le 1, \operatorname{spt}|g| \subset W \\ g \text{ smooth}}} \int u \operatorname{div} g \, d\mathcal{L}^n , \quad W \text{ open} \subset U .$$

The left side here is more usually denoted $\int_W |Du|$. Indeed if f is any non-negative Borel measurable function function on U, then $\int f \, d|Du|$ is more usually denoted simply by $\int f|Du|$ ($\equiv \int f|Du| \, d\mathcal{L}^n$ in case $u \in W^{1,1}_{loc}(U)$).

We shall henceforth adopt this notation.

There are a number of important results about BV functions which can be obtained by mollification. We let $\phi_\sigma(x) = \sigma^{-n}\phi(\frac{x}{\sigma})$, where ϕ is a symmetric mollifier (so that $\phi \in C_c^\infty(\mathbb{R}^n)$, $\phi \ge 0$, $\operatorname{spt}\phi \subset B_1(0)$, $\int_{\mathbb{R}^n}\phi = 1$, and $\phi(x) = \phi(-x)$), and for $u \in L^1_{loc}(U)$ let $u^{(\sigma)} = \phi_\sigma * u$ be the mollified functions, where we set $\tilde{u} = u$ on U_σ, $\tilde{u} = 0$ outside U_σ, $U_\sigma = \{x \in U : \operatorname{dist}(x,\partial U) > \sigma\}$. A key result concerning mollification is then as follows:

6.2 LEMMA *If* $u \in BV_{loc}(U)$, *then* $u^{(\sigma)} \to u$ *in* $L^1_{loc}(U)$ *and* $|Du^{(\sigma)}| \to |Du|$ *in the sense of Radon measures in* U (see 4.4) *as* $\sigma \downarrow 0$.

Proof The convergence of $u^{(\sigma)}$ to u in $L^1_{loc}(U)$ is standard. Thus it remains to prove

$$\lim_{\sigma\downarrow 0}\int f|Du^{(\sigma)}| = \int f|Du| \tag{1}$$

for each $f \in C_c^0(U)$, $f \ge 0$. In fact by definition of $|Du|$ it is rather easy to prove that

$$\int f|Du| \le \liminf_{\sigma\downarrow 0}\int f|Du^{(\sigma)}| ,$$

so we only have to check

$$\limsup_{\sigma\downarrow 0}\int f|Du^{(\sigma)}| \le \int f|Du| \tag{2}$$

for each $f \in C_c^0(U)$, $f \ge 0$.

This is achieved as follows: First note that

$$\int f|Du^{(\sigma)}| = \sup_{|g|\le f,\, g \text{ smooth}}\int g\cdot \operatorname{grad} u^{(\sigma)}\, dL^n . \tag{3}$$

On the other hand for fixed g with g smooth and $|g| \le f$, and for

$\sigma < \mathrm{dist}(\mathrm{spt}\, f, \partial U)$, we have

$$\begin{aligned}\int g \cdot \mathrm{grad}\, u^{(\sigma)}\, dL^n &= -\int u^{(\sigma)} \mathrm{div}\, g \;\; dL^n \\ &= -\int \phi_\sigma * u \;\mathrm{div}\, g \; dL^n \\ &= -\int u(\phi_\sigma * \mathrm{div}\, g)\, dL^n \\ &= -\int u \;\mathrm{div}(\phi_\sigma * g)\, dL^n \; .\end{aligned}$$

On the other hand by definition of $|Du|$, the right side here is

$$\le \int_{W_\sigma} (f+\varepsilon(\sigma))\, |Du|$$

where $\varepsilon(\sigma) \downarrow 0$, where $W = \mathrm{spt}\, f$, $W_\sigma = \{x \in U : \mathrm{dist}(x,W) < \sigma\}$, because

$$\begin{aligned}|\phi_\sigma * g| &\equiv |(\phi_\sigma * g^1, \ldots, \phi_\sigma * g^n)| \\ &\le \phi_\sigma * |g| \le \phi_\sigma * f\end{aligned}$$

and because $\phi_\sigma * f \to f$ uniformly in W_{σ_0} as $\sigma \downarrow 0$, where $\sigma_0 < \mathrm{dist}(W, \partial U)$. Thus (2) follows from (3).

6.3 THEOREM (Compactness Theorem for BV function)

If $\{u_k\}$ *is a sequence of* $BV_{loc}(U)$ *functions satisfying*

$$\sup_{k\ge 1}\left(\|u_k\|_{L^1(W)} + \int_W |Du_k|\right) < \infty$$

for each $W \subset\subset U$, *then there is a subsequence* $\{u_{k'}\} \subset \{u_k\}$ *and a* $BV_{loc}(U)$ *function* u *such that* $u_{k'} \to u$ *in* $L^1_{loc}(U)$ *and*

$$\int_W |Du| \le \liminf \int_W |Du_{k'}| \qquad \forall\, W \subset\subset U \; .$$

Proof By virtue of the previous lemma, in order to prove $u_{k'} \to u$ in

$L^1_{loc}(U)$ for some subsequence $\{u_{k'}\}$, it is enough to prove that the sets

$$\left\{u\in C^\infty(U) : \int_W (|u|+|Du|)\,dL^n \le c(W)\right\},\quad W\subset\subset U\ ,$$

(for given constants $c(W)<\infty$) are precompact in $L^1_{loc}(U)$. For the simple proof of this (involving mollification and Arzela's theorem) see for example [GT, Theorem 7.22].

Finally the fact that $\int_W |Du| \le \liminf \int_W |Du_{k'}|$ is a direct consequence of the definition of $|Du|$, $|Du_{k'}|$.

Next we have the Poincaré inequality for BV functions.

6.4 LEMMA *Suppose* U *is bounded, open and convex,* $u\in BV_{loc}(\mathbb{R}^n)$ *with* $\operatorname{spt} u\subset \bar U$. *Then for any* $\theta\in(0,1)$ *and any* $\beta\in\mathbb{R}$ *with*

(*) $\min\{L^n\{x\in U : u(x)\ge\beta\}, L^n\{x\in U : u(x)\le\beta\}\}\ge \theta\, L^n(U)$.

we have

$$\int_U |u-\beta|\ dL^n \le c\int_U |Du|\ ,$$

where $c=c(\theta,U)$.

Proof Let β,θ be as in (*) and choose convex $W\subset U$ such that

(1) $$|Du|(\partial W)=0\ ,\quad \int_W |u-\beta|\ dL^n \ge \frac{1}{2}\int_U |u-\beta|\,dL^n$$

and such that (*) holds with W in place of U and $\theta/2$ in place of θ. (For example we may take $W=\{x\in U : \operatorname{dist}(x,\partial U)>\eta\}$ with η small.)

Letting $u^{(\sigma)}$ denote the mollified functions corresponding to u, note that for sufficiently small σ we must then have (*) with $u^{(\sigma)}$ in place of u, $\theta/4$ in place of θ, $\beta^{(\sigma)}\to\beta$ in place of β, and W in place of U. Hence by the usual Poincaré inequality for smooth functions (see e.g. [GT]) we have

$$\int_W |u^{(\sigma)}-\beta^{(\sigma)}|\,dL^n \le c \int_W |Du^{(\sigma)}|\, dL^n ,$$

$c = c(n,\theta,W)$, for all sufficiently small σ . The required inequality now follows by letting $\sigma \downarrow 0$ and using (1) above together with 6.2.

6.5 LEMMA *Suppose* U *is bounded, open and convex,* $u \in BV_{loc}(\mathbb{R}^n)$ *with* $\operatorname{spt} u \subset \bar{U}$. *Then*

$$\int_{\mathbb{R}^n} |Du| \left(\equiv \int_{\bar{U}} |Du| \right) \le c\left(\int_U |Du| + \int_U |u|\, dL^n \right) ,$$

where $c = c(U)$.

6.6 REMARK Note that by combining this with the Poincaré inequality 6.4, we conclude

$$\int_{\mathbb{R}^n} |D(u-\beta\chi_U)| \le c \int_U |Du| ,$$

$c = c(\theta,U)$, whenever β is as in $(*)$ of 6.4.

Proof of 6.5 Let $U_\delta = \{x \in U : \operatorname{dist}(x,\partial U) > \delta\}$ and (for δ small) let ϕ_δ be a $C_c^\infty(\mathbb{R}^n)$ function satisfying

$$(1) \qquad \phi_\delta = \begin{cases} 1 \text{ in } U_\delta \\ 0 \text{ in } \mathbb{R}^n \sim U_{\delta/2} \end{cases}$$

$$(2) \qquad 0 \le \phi_\delta \le 1 \quad \text{in} \quad \mathbb{R}^n ,$$

and (for a given point $a \in U$)

$$(3) \qquad |D\phi_\delta(x)| \le -c(x-a)\cdot D\phi_\delta(x) , \quad x \in U ,$$

where $c = c(U,a)$ is independent of δ . (One easily checks that such ϕ_δ

exist, for sufficiently small δ, because of the convexity of U.)

Now by definition of $|Dw|$ for $BV_{loc}(\mathbb{R}^n)$ functions w, we have

$$(4) \qquad \int_{\mathbb{R}^n} |D(\phi_\delta u)| \le \int_{\mathbb{R}^n} |D\phi_\delta|\ |u|\ dL^n + \int_{\mathbb{R}^n} \phi_\delta |Du|$$

and by (3)

$$(5) \qquad c^{-1} \int_{\mathbb{R}^n} |D\phi_\delta|\ |u|\ dL^n \le -\int (x-a) \cdot D\phi_\delta\ |u|\ dL^n$$

$$\equiv -\int (|u|\ \operatorname{div}((x-a)\phi_\delta) + n|u|\phi_\delta)\ dL^n$$

$$\le c\left(\int_U |D|u|| + \int_{\mathbb{R}^n} |u|\ dL^n \right)$$

(by definition of $|D|u||$)

$$\le c\left(\int_U |Du| + \int_{\mathbb{R}^n} |u|\ dL^n \right)$$

(because $|D|u|| \le |Du|$ by virtue of 6.2 and the fact that $|D|u|| \le \liminf_{\sigma \downarrow 0} |D|u^{(\sigma)}||$).

Finally, to complete the proof of 6.5, we note that (using the definition of $|Dw|$ for the $BV_{loc}(\mathbb{R}^n)$ functions $w = u, \phi_\delta u$, together with the fact that $\phi_\delta u \to u$ in $L^1(\mathbb{R}^n)$)

$$\int_{\mathbb{R}^n} |Du| \le \liminf_{\delta \downarrow 0} \int_{\mathbb{R}^n} |D(\phi_\delta u)| .$$

Then 6.5 follows from (4), (5).

7. SUBMANIFOLDS OF $\mathbb{R}^{n+k}$

Let M denote an n-dimensional C^r submanifold of $\mathbb{R}^{n+k}$, $0 \le k$, $r \ge 1$. By this we mean that for each $y \in M$ there are open sets $U, V \subset \mathbb{R}^{n+k}$ with

$y \in U$, $0 \in V$ and a C^r diffeomorphism $\phi : U \to V$ such that $\phi(y) = 0$ and

$$\phi(M \cap U) = W = V \cap \mathbb{R}^n .$$

(Here and subsequently we identify $\mathbb{R}^n$ with the subspace of $\mathbb{R}^{n+k}$ consisting of all points $x = (x^1,\ldots,x^{n+k})$ such that $x^j = 0$, $j = n+1,\ldots,n+k$.)

In particular we have *local representations*

$$\psi : W \to \mathbb{R}^{n+k} , \quad \psi(W) = M \cap V , \quad \psi(0) = y$$

such that $\frac{\partial\psi}{\partial x^1}(0) , \ldots , \frac{\partial\psi}{\partial x^n}(0)$ are linearly independent vectors in $\mathbb{R}^{n+k}$. (For example we can take $\psi = \phi^{-1}|W$.) The tangent space T_yM of M at y is the subspace of $\mathbb{R}^{n+k}$ consisting of those $\tau \in \mathbb{R}^{n+k}$ such that

$$\tau = \dot{\gamma}(0) \text{ for some } C^1 \text{ curve } \gamma : (-1,1) \to \mathbb{R}^{n+k}, \gamma(-1,1) \subset M , \gamma(0) = y .$$

One readily checks that T_yM has a basis $\frac{\partial\psi}{\partial x^1}(0) , \ldots , \frac{\partial\psi}{\partial x^n}(0)$ for a local representation ψ as above.

A function $f : M \to \mathbb{R}^N$ $(N\geq 1)$ is said to be C^ℓ $(\ell\leq r)$ on M if f is the restriction to M of a C^ℓ function $\bar{f} : U \to \mathbb{R}^N$, where U is an open set in $\mathbb{R}^{n+k}$ such that $M \subset U$.

Given $\tau \in T_yM$ the directional derivative $D_\tau f \in \mathbb{R}^N$ is defined by $\frac{d}{dt} f(\gamma(t))\Big|_{t=0}$ for any C^1 curve $\gamma : (-1,1) \to M$ with $\gamma(0) = y$, $\dot{\gamma}(0) = \tau$. Of course it is easy to see that this definition is independent of the particular curve γ we use to represent τ . The induced linear map $df_y : T_yM \to \mathbb{R}^N$ is defined by $df_y(\tau) = D_\tau f$, $\tau \in T_yM$. (Evidently df_y is exactly characterized by being the "best linear approximation" to f at y in the obvious sense.)

In case f is real-valued (i.e. $N=1$) then we define the gradient $\nabla^M f$ of f by

$$\nabla^M f(y) = \sum_{j=1}^{n} (D_{\tau_j} f)\tau_j \ , \quad y \in T_y M \ ,$$

$\tau_1,\ldots,\tau_n$ any orthonormal basis for $T_y M$. If we let $\nabla_j^M f \equiv e_j \cdot \nabla^M f$ (e_j = j-th standard basis vector in $\mathbb{R}^{n+k}$, $j = 1,\ldots,n+k$) then

$$\nabla^M f(y) = \sum_{j=1}^{n+k} \nabla_j^M f(y) e_j \ .$$

If f is the restriction to M of a $C^1(U)$ function $\bar{f}$, where U is an open subset of $\mathbb{R}^{n+k}$ containing M, then

$$\nabla^M f(y) = (\operatorname{grad}_{\mathbb{R}^{n+k}} \bar{f}(y))^T \ , \quad y \in M \ ,$$

where $\operatorname{grad}_{\mathbb{R}^{n+k}} \bar{f}$ is the usual $\mathbb{R}^{n+k}$ gradient $(D_1\bar{f},\ldots,D_{n+k}\bar{f})$ on U, and where $(\ )^T$ means orthogonal projection of $\mathbb{R}^{n+k}$ onto $T_y M$.

Now given a vector function ("vector field") $X = (X^1,\ldots,X^{n+k}) : M \to \mathbb{R}^{n+k}$ with $X^j \in C^1(M)$, $j = 1,\ldots,n+k$, we define

$$\operatorname{div}_M X = \sum_{j=1}^{n+k} \nabla_j^M X^j$$

on M. (Notice that we do *not* require $X_y \in T_y M$.) Then, at $y \in M$, we have

$$\begin{aligned} \operatorname{div}_M X &= \sum_{j=1}^{n+k} e_j \cdot (\nabla^M X^j) \\ &= \sum_{j=1}^{n+k} e_j \cdot \left(\sum_{i=1}^{n} (D_{\tau_j} X^j)\tau_i \right) , \end{aligned}$$

so that (since $X = \sum_{j=1}^{n+k} X^j e_j$)

$$\operatorname{div}_M X = \sum_{i=1}^{n} (D_{\tau_i} X) \cdot \tau_i \ ,$$

where $\tau_1,\ldots,\tau_n$ is any orthonormal basis for T_yM .

The *divergence theorem* states that if the closure $\bar{M}$ of M is a smooth compact manifold with boundary $\partial M = \bar{M} \sim M$, and if $X_y \in T_yM$ $\forall\, y \in M$, then

7.1 $$\int_M \operatorname{div}_M X \, dH^n = -\int_{\partial M} X\cdot\eta \, dH^{n-1}$$

where η is the inward pointing unit co-normal of ∂M ; that is, $|\eta| = 1$, η is normal to ∂M , tangent to M , and points into M at each point of ∂M .

7.2 REMARKS

(1) M need *not* be orientable here.

(2) In general the closure $\bar{M}$ of M will not be a nice manifold with boundary; indeed it can certainly happen that $H^n(\bar{M}\sim M) > 0$. (For example consider the case when $M = \bigcup_{k=1}^{\infty} \{(x,y) \in \mathbb{R}^2 : y = x^2/k\} \sim \{0\}$. M is a C^r 1-dimensional submanifold of $\mathbb{R}^2$ $\forall\, r$ in the sense of the above definitions, but $\bar{M} \sim M$ is the whole x-coordinate axis.) Nevertheless in the general case we still have (in place of 7.1)

$$\int_M \operatorname{div}_M X = 0$$

provided support $X \cap M$ *is a compact subset of* M and $X_y \in T_yM$ $\forall\, y \in M$.

In case M is at least C^2 we define *the second fundamental form* of M at y to be the bilinear form

$$B_y : T_yM \times T_yM \to (T_yM)^{\perp}$$

such that

7.3 $$B_y(\tau,\eta) = -\sum_{\alpha=1}^{k} (\eta\cdot D_\tau \nu^\alpha)\nu^\alpha\big|_y \ , \quad \tau,\eta \in T_yM \ ,$$

where $\nu^1, \ldots, \nu^k$ are (locally defined, near y) vector fields with $\nu^\alpha(z) \cdot \nu^\beta(z) = \delta_{\alpha\beta}$ and $\nu^\alpha(z) \in (T_z M)^\perp$ for every z in some neighbourhood of y. The geometric significance of B is as follows: If $\tau \in T_y M$ with $|\tau| = 1$ and $\gamma : (-1,1) \to \mathbb{R}^{n+k}$ is a C^2 curve with $\gamma(0) = y$, $\gamma(-1,1) \subset M$, and $\dot\gamma(0) = \tau$, then

$$B_y(\tau,\tau) = (\ddot\gamma(0))^\perp ,$$

which is just the normal component (relative to M) of the *curvature* of γ at 0, γ being considered as an ordinary space-curve in $\mathbb{R}^{n+k}$. (Thus $B_y(\tau,\tau)$ measures the "normal curvature" of M in the direction τ.) To check this, simply note that $\nu^\alpha(\gamma(t)) \cdot \dot\gamma(t) \equiv 0$, $|t| < 1$, because $\dot\gamma(t) \in T_{\gamma(t)} M$ and $\nu^\alpha(\gamma(t)) \in (T_{\gamma(t)} M)^\perp$. Differentiating this relation with respect to t, we get

(after setting $t = 0$)

$$\nu^\alpha(y) \cdot \ddot\gamma(0) = -(D_\tau \nu^\alpha) \cdot \tau$$

and hence (multiplying by $\nu^\alpha(y)$ and summing over α) we have

$$(\ddot\gamma(0))^\perp = -\sum_{\alpha=1}^{k} (\tau \cdot D_\tau \nu^\alpha)\nu^\alpha(y)$$

$$= B_y(\tau,\tau)$$

as required. (Note that the parameter t here need *not* be arc-length for γ; it suffices that $\dot\gamma(0) = \tau$, $|\tau| = 1$.) More generally, by a similar argument, if $\tau, \eta \in T_y M$ and if $\phi : U \to \mathbb{R}^{n+k}$ is a C^2 mapping of a neighbourhood U of 0 in $\mathbb{R}^2$ such that $\phi(U) \subset M$, $\phi(0) = y$, $\frac{\partial\phi}{\partial x^1}(0,0) = \tau$, $\frac{\partial\phi}{\partial x^2}(0,0) = \eta$, then

$$B_y(\tau,\eta) = \left(\frac{\partial^2\phi}{\partial x^1 \partial x^2}(0,0)\right)^\perp .$$

In particular $B_y(\tau,\eta) = B_y(\eta,\tau)$; that is B_y is a *symmetric* bilinear form with values in $(T_yM)^\perp$.

We define the mean curvature vector $\underline{\underline{H}}$ of M at y to be trace B_y ; thus

7.4 $$\underline{\underline{H}}(y) = \sum_{i=1}^{n} B_y(\tau_i,\tau_i) \in (T_yM)^\perp ,$$

where $\tau_1, \ldots, \tau_n$ is any orthonormal basis for T_yM . Notice that then (if $\nu^1,\ldots,\nu^k$ are as above)

$$\underline{\underline{H}}(y) = -\sum_{\alpha=1}^{k}\sum_{i=1}^{n} (\tau_i \cdot D_{\tau_i}\nu^\alpha)\nu^\alpha(y)$$

so that

7.5 $$\underline{\underline{H}} = -\sum_{\alpha=1}^{k} (\mathrm{div}_M \nu^\alpha)\nu^\alpha$$

near y .

Returning for a moment to 7.1 (in case $\bar{M}$ is a compact C^2 manifold with smooth $(n-1)$-dimensional boundary $\partial M = \bar{M} \sim M$) it is interesting to compute $\int_M \mathrm{div}_M X$ in case the condition $X_y \in T_yM$ is *dropped*. To compute this, we decompose X into its tangent and normal parts:

$$X = X^\top + X^\perp ,$$

where (at least locally, in the notation introduced above)

$$X^\perp = \sum_{\alpha=1}^{k} (\nu^\alpha \cdot X)\nu^\alpha .$$

Then we have (near y)

$$\mathrm{div}_M X^\perp = \sum_{\alpha=1}^{k} (\nu^\alpha \cdot X)\,\mathrm{div}\,\nu^\alpha ,$$

so that by 7.5

7.5' $$\mathrm{div}_M X^{\perp} = - X \cdot \underline{\underline{H}}$$

at each point of M. On the other hand $\int_M \mathrm{div}_M X^T = - \int_{\partial M} X \cdot \eta$ by 7.1. Hence, since $\mathrm{div}_M X = \mathrm{div}_M X^T + \mathrm{div}_M X^{\perp}$, we obtain

7.6 $$\int_M \mathrm{div}_M X \, d\mathcal{H}^n = - \int_M X \cdot \underline{\underline{H}} \, d\mathcal{H}^n - \int_{\partial M} X \cdot \eta \, d\mathcal{H}^{n-1} .$$

§8. THE AREA FORMULA

Recall that if λ is a linear map $\mathbb{R}^n \to \mathbb{R}^n$ and $A \subset \mathbb{R}^n$, then $\mathcal{L}^n(\lambda(A)) = |\det \lambda| \, \mathcal{L}^n(A)$. More generally if $\lambda : \mathbb{R}^n \to \mathbb{R}^N$, $N \geq n$, then $\lambda(\mathbb{R}^n) \subset F$ where F is a n-dimensional subspace of $\mathbb{R}^N$, and hence choosing an orthogonal transformation q of $\mathbb{R}^N$ such that $q(F) = \mathbb{R}^n$, we see that $q \circ \lambda : \mathbb{R}^n \to \mathbb{R}^n$ and hence $\mathcal{L}^n(q\lambda(A)) = |\det(q\lambda)| \mathcal{L}^n(A)$ for $A \subset \mathbb{R}^n$. One readily checks, since q is orthogonal, that $|\det(q\lambda)| = \sqrt{\det \lambda^* \circ \lambda}$, where $\lambda^* : \mathbb{R}^N \to \mathbb{R}^n$ is the adjoint of λ. Since $\mathcal{H}^n(q(B)) = \mathcal{H}^n(B)$ (by definition of $\mathcal{H}^n$) we have by Theorem 2.8 that $\mathcal{L}^n(q\lambda(A)) = \mathcal{H}^n(q\lambda(A)) = \mathcal{H}^n(\lambda(A))$, and hence we obtain the area formula

8.1 $$\mathcal{H}^n(\lambda(A)) = \sqrt{\det \lambda^* \circ \lambda} \; \mathcal{H}^n(A) , \quad A \subset \mathbb{R}^n ,$$

whenever λ is a linear map $\mathbb{R}^n \to \mathbb{R}^N$, $n \leq N$.

More generally given a 1:1 C^1 map $f : M \to \mathbb{R}^N$ (M an n-dimensional C^1 submanifold of $\mathbb{R}^{n+k}$) we have, by an approximation argument based on the linear case 8.1 (see [HR] or [FH1] for details) that

8.2 $$\mathcal{H}^n(f(A)) = \int_A Jf \, d\mathcal{H}^n \quad \forall \ \mathcal{H}^n\text{-measurable set } A \subset M ,$$

where Jf is the Jacobian of f (or area magnification factor of f) defined by

8.3 $$Jf(y) = \sqrt{\det(df_y)^* \circ (df_y)} \quad .$$

Here $df_y : T_y M \to \mathbb{R}^N$ is the induced linear map described in §7, and $(df_y)^* : \mathbb{R}^N \to T_y M$ denotes the adjoint transformation.

If f is not 1:1 we have the general area formula (which actually follows quite easily from 8.2)

8.4 $$\int_{\mathbb{R}^N} \mathcal{H}^0(f^{-1}(y) \cap A)\, d\mathcal{H}^n(y) = \int_A Jf \, d\mathcal{H}^n \;, \quad \forall\ \mathcal{H}^n\text{-measurable } A \subset M \;,$$

where $\mathcal{H}^0$ is 0-dimensional Hausdorff measure i.e. "counting measure". (Thus $\mathcal{H}^0(B) = 0$ if $B = \emptyset$, $\mathcal{H}^0(B)$ = the number of elements of the set B if B is a finite non-empty set, and $\mathcal{H}^0(B) = \infty$ if B is not finite). More generally still, if g is a non-negative $\mathcal{H}^n$-measurable function on M, then

8.5 $$\int_{\mathbb{R}^N} \int_{f^{-1}(y)} g \, d\mathcal{H}^0 \, d\mathcal{H}^n(y) = \int_M (Jf)\, g \, d\mathcal{H}^n \;.$$

This follows directly from 8.4 if we approximate g by simple functions.

8.6 EXAMPLES

(1) *Space curves.* Using the above area formula we first check that $\mathcal{H}^1$-measure agrees with the usual arc-length measure for C^1 curves in $\mathbb{R}^n$. In fact if $\gamma : [a,b] \to \mathbb{R}^n$ is a 1:1 C^1 map then the Jacobian is just $\sqrt{|\dot\gamma|^2} = |\dot\gamma|$, so that 8.2 gives

$$\mathcal{H}^1(\gamma(A)) = \int_A |\dot\gamma| \, dL^1$$

as required.

(2) *Submanifolds of* $\mathbb{R}^{n+k}$. If M is any n-dimensional C^1 manifold of $\mathbb{R}^{n+k}$, we want to check that $\mathcal{H}^n$ agrees with the usual n-dimensional volume measure on M. It is enough to check this in a region where a local coodinate representation as in §7 applies. If

$$\psi : W \to \mathbb{R}^{n+k}, \quad \psi(W) = M \cap U$$

is a local representation for M as in §7 then the usual definition of the n-dimensional volume of a Borel set $A \subset M \cap U$ is

$$\mu(A) = \int_{\tilde{A}} \sqrt{g}\, d\mathcal{L}^n ,$$

where $\tilde{A} = \psi^{-1}(A)$ and $g = \det(g_{ij})$, $g_{ij} = \frac{\partial\psi}{\partial x^i} \cdot \frac{\partial\psi}{\partial x^j}$, $i,j = 1,\ldots,n$. However one easily checks that then $\sqrt{g}$ is precisely $J\psi$, the Jacobian of $\psi : W \to \mathbb{R}^{n+k}$, defined as above. Hence we have by the area formula 8.2 that $\int_{\tilde{A}} \sqrt{g}\, d\mathcal{L}^n = \mathcal{H}^n(\psi(\tilde{A})) = \mathcal{H}^n(A)$, so that $\mu(A) = \mathcal{H}^n(A)$.

(3) *n-dimensional graphs in* $\mathbb{R}^{n+1}$. If Ω is a domain in $\mathbb{R}^n$ and if $M = \operatorname{graph} u$, where $u \in C^1(\Omega)$, then M is globally represented by the map $\psi : x \mapsto (x,u(x))$; in this case $J\psi(x) \equiv \sqrt{\det\left(\frac{\partial\psi}{\partial x^i} \cdot \frac{\partial\psi}{\partial x^j}\right)}$

$$\equiv \sqrt{\det(\delta_{ij} + D_i u D_j u)} = \sqrt{1 + |Du|^2} ,$$

so $\mathcal{H}^n(M) = \int_\Omega \sqrt{1+|Du|^2}\, dx$ (by (2) above).

§9. FIRST AND SECOND VARIATION FORMULAE

Suppose that M is an n-dimensional C^1 submanifold of $\mathbb{R}^{n+k}$ and let U be an open subset of $\mathbb{R}^{n+k}$ such that $U \cap M \neq \emptyset$ and $\mathcal{H}^n(K \cap M) < \infty$ for each compact $K \subset U$. Also, let $\{\phi_t\}_{0 \le t \le 1}$ be a 1-parameter family of diffeomorphisms $U \to U$ such that

$$9.1 \quad \begin{cases} (1) \quad \phi(t,x)\,(\equiv \phi_t(x)) \quad \text{is a } C^2 \text{ map}: (-1,1)\times U \to U \\ (2) \quad \phi_0(x) \equiv x\,, \quad x \in U \\ (3) \quad \phi_t(x) \equiv x \quad \forall\, t \in (-1,1)\,, \quad x \in U \sim K\,, \quad \text{where } K \subset U \end{cases}$$

is a compact subset of U.

Also, let X, Z denote the initial velocity and acceleration vectors for ϕ_t : thus $X_x = \left.\dfrac{\partial\phi(t,x)}{\partial t}\right|_{t=0}$, $Z_x = \left.\dfrac{\partial^2\phi(t,x)}{\partial t^2}\right|_{t=0}$.

Then

$$9.2 \qquad \phi_t(x) = x + tX_x + \frac{t^2}{2} Z_x + O(t^3)$$

and X, Z have supports which are compact subsets of U. Let $M_t = \phi_t(M \cap K)$ (K as in 9.1 (3)); thus M_t is a 1-parameter family of manifolds such that $M_0 = M \cap K$ and M_t agrees with M outside some compact subset of U. We want to compute $\left.\dfrac{d}{dt} \mathcal{H}^n(M_t)\right|_{t=0}$ and $\left.\dfrac{d^2}{dt^2} \mathcal{H}^n(M_t)\right|_{t=0}$ (i.e. the "first and second variation" of M). The area formula is particularly useful here because it gives (with K as in 9.1 (3))

$$\mathcal{H}^n(\phi_t(M\cap K)) = \int_{M\cap K} J\psi_t \, d\mathcal{H}^n\,, \quad \psi_t = \phi_t | M \cap U\,,$$

and hence to compute the first and second variation we can differentiate under the integral. Thus the computation reduces to calculation of

$$\left.\frac{\partial}{\partial t} J\psi_t\right|_{t=0} \quad \text{and} \quad \left.\frac{\partial^2}{\partial t^2} J\psi_t\right|_{t=0} .$$

To calculate we first want to get a manageable expression for $J\psi_t$. First note that (for fixed t)

$$\begin{aligned} d\psi_t\big|_x(\tau) &= D_\tau \psi_t \qquad (\tau \in T_xM) \\ &= \tau + t\, D_\tau X + \frac{t^2}{2} D_\tau Z + O(t^3) \qquad \text{(by 9.2)} . \end{aligned}$$

Hence, relative to the bases $\tau_1,\ldots,\tau_n$ for T_xM and $e_1,\ldots,e_{n+k}$ for $\mathbb{R}^{n+k}$, the map $d\psi_t|_x : T_xM \to \mathbb{R}^{n+k}$, has matrix

$$a_{\ell i} = \tau_i^\ell + t\, D_{\tau_i} X^\ell + \frac{t^2}{2} D_{\tau_i} Z^\ell + O(t^3)$$

for $i = 1,\ldots,n$, $\ell = 1,\ldots,n+k$. Then $(d\psi_t|_x)^* \circ (d\psi_t|_x)$ has matrix $\left(\sum_{\ell=1}^{n+k} a_{\ell i}a_{\ell j}\right)_{i,j=1,\ldots,n} \equiv (b_{ij})$, where

$$\begin{aligned} b_{ij} = \delta_{ij} &+ t(\tau_i\cdot D_{\tau_j}X+\tau_j\cdot D_{\tau_i}X) \\ &+ t^2(\tfrac{1}{2}(\tau_i\cdot D_{\tau_j}Z+\tau_j\cdot D_{\tau_i}Z) + (D_{\tau_i}X)\cdot(D_{\tau_j}X)) \\ &+ O(t^3), \end{aligned}$$

so that (by the general formula $\det(I+tA+t^2B) = 1 + t\,\text{trace}\,A + t^2(\text{trace}\,B + \frac{1}{2}(\text{trace}\,A)^2 - \frac{1}{2}\,\text{trace}\,(A^2)) + O(t^3))$ we have

$$\begin{aligned} (J\psi_t)^2 &= 1 + 2t\,\text{div}_M X + t^2(\text{div}_M Z + \sum_{i=1}^n |D_{\tau_i}X|^2 \\ &\qquad + 2(\text{div}_M X)^2 - \frac{1}{2}\sum_{i,j=1}^n (\tau_i\cdot D_{\tau_j}X + \tau_j\cdot D_{\tau_i}X)^2 + O(t^3) \\ &= 1 + 2t\,\text{div}_M X + t^2(\text{div}_M Z + \sum_{i=1}^n |(D_{\tau_i}X)^\perp|^2 \\ &\qquad + 2(\text{div}_M X)^2 - \sum_{i,j=1}^n (\tau_i\cdot D_{\tau_j}X)(\tau_j\cdot D_{\tau_i}X)) + O(t^3), \end{aligned}$$

where $(D_{\tau_i}X)^\perp$ ($\equiv$ normal part of $D_{\tau_i}X) = D_{\tau_i}X - \sum_{j=1}^n (\tau_j\cdot D_{\tau_i}X)\tau_j$.

Using $\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + O(x^3)$, we thus get

$$\begin{aligned} J\psi_t = 1 + t\,\text{div}_M X + \frac{t^2}{2}(\text{div}_M Z + (\text{div}_M X)^2 + \sum_{i=1}^n |(D_{\tau_i}X)^\perp|^2 \\ - \sum_{i,j=1}^n (\tau_i\cdot D_{\tau_j}X)(\tau_j\cdot D_{\tau_i}X)) + O(t^3). \end{aligned}$$

Thus the area formula immediately yields the first and second variation formulae:

9.3 $$\frac{d}{dt} \mathcal{H}^n(M_t)\Big|_{t=0} = \int_M \operatorname{div}_M X \, d\mathcal{H}^n$$

and

9.4 $$\frac{d^2}{dt^2} \mathcal{H}^n(M_t)\Big|_{t=0} = \int_M (\operatorname{div}_M Z + (\operatorname{div}_M X)^2 + \sum_{i=1}^n |(D_{\tau_i} X)^\perp|^2 - \sum_{i,j=1}^n (\tau_i \cdot D_{\tau_j} X)(\tau_j \cdot D_{\tau_i} X)) .$$

We shall use the terminology that M *is stationary in* U if $\mathcal{H}^n(M \cap K) < \infty$ for each compact $K \subset U$ and if $\frac{d}{dt} \mathcal{H}^n(M_t)\Big|_{t=0} = 0$ whenever $M_t = \phi_t(M \cap K)$, K, ϕ_t as in 9.1. Thus in view of 9.3 we see that M is *stationary in* U *if and only if* $\int_M \operatorname{div}_M X \, d\mathcal{H}^n = 0$ whenever X is C^1 on U with support X a compact subset of U .

In view of 7.6 we also have the following

9.5 LEMMA

(1) *If* M *is a* C^2 *submanifold of* $\mathbb{R}^{n+k}$ *and* $\bar{M}$ *is a* C^2 *submanifold with smooth* $(n-1)$*-dimensional boundary* $\partial M = \bar{M} \sim M$, *then* M *is stationary in* U *if and only if* $\underline{\underline{H}} \equiv 0$ *on* $M \cap U$ *and* $\partial M \cap U = \emptyset$.

(2) *Generally, if* M *is an arbitrary* C^2 *submanifold of* $\mathbb{R}^{n+k}$ *and* $\bar{U} \cap M$ *is a compact subset of* M , *then* M *is stationary in* U *if and only if* $\underline{\underline{H}} \equiv 0$ *on* $M \cap U$.

(In both parts (1), (2) above $\underline{\underline{H}}$ denotes the mean curvature vector of M .)

For later reference we also want to mention an important modification of the idea that M be stationary in U , U open in $\mathbb{R}^{n+k}$. Namely, suppose N is a C^2 $(n+k_1)$-dimensional submanifold of $\mathbb{R}^{n+k}$, $0 \le k_1 \le k$, and *suppose* U *is an open subset of* N *and* $M \subset N$. Then *we say that* M *is stationary in* U *if* 9.3 *holds whenever* $X_y \in T_y N$ $\forall\, y \in M$. This is evidently equivalent to the requirement that $\frac{d}{dt} H^n(\phi_t(M \cap K))\Big|_{t=0} = 0$ whenever ϕ_t satisfies the conditions 9.1 (bearing in mind that U is required now to be an open subset of N rather than an open subset of $\mathbb{R}^{n+k}$ as before). If we let $\nu^1,\ldots,\nu^k$ be an orthonormal family (defined locally near a point $y \in M$) of vector fields normal to M, such that $\nu^1,\ldots,\nu^{k_1}$ are tangent to N and $\nu^{k_1+1},\ldots,\nu^k$ are normal to N, then for any vector field X on M we can write $X = X^{(1)} + X^{(2)}$, where $X_z^{(1)} \in T_z N$ and $X^{(2)} = \sum_{j=k_1+1}^{n} (\nu^j \cdot X)\nu^j$ (= part of X normal to N). Then if $\tau_1,\ldots,\tau_n$ is any orthonormal basis for $T_y M$, we have

$$\begin{aligned}\operatorname{div}_M X &= \operatorname{div}_M X^{(1)} + \sum_{j=k_1+1}^{n} (\nu^j \cdot X)\operatorname{div}_M \nu^j \\ &\equiv \operatorname{div}_M X^{(1)} + \sum_{i=1}^{n} X \cdot \bar{B}_y(\tau_i,\tau_i) ,\end{aligned}$$

where $\bar{B}_y$ is the second fundamental form of N at y,

Thus we conclude

9.6 LEMMA *If* N *is an* $(n+k_1)$*-dimensional submanifold of* $\mathbb{R}^{n+k}$, *if* $M \subset N$ *and if* U *is an open subset of* N *such that* $H^n(M \cap K) < \infty$ *whenever* K *is a compact subset of* U, *then* M *is stationary in* U *if and only if*

$$\int_M \operatorname{div}_M X = - \int_M \bar{\underline{\underline{H}}}_M \cdot X$$

for each C^1 *vector field* X *with compact support contained in* U; *here* $\bar{\underline{\underline{H}}}_M\big|_y = \sum_{i=1}^{n} \bar{B}_y(\tau_i,\tau_i)$, $y \in M$, *where* $\bar{B}_y$ *denotes the second fundamental form of* N *at* y *and* $\tau_1,\ldots,\tau_n$ *is any orthonormal basis of* $T_y M$.

Finally, we shall need later the following important fact about second variation formula 9.4.

9.7 LEMMA *If* M *is* C^2 *, stationary in* U *,* U *open in* $\mathbb{R}^{n+k}$ *with* $(\bar{M} \sim M) \cap U = \emptyset$ *, and if* X *as in* 9.4 *has compact support in* U *with* $X_y \in (T_y M)^{\perp}$ $\forall\, y \in M$ *, then* 9.4 *says*

$$\frac{d^2}{dt^2} \mathcal{H}^n(M_t)\Big|_{t=0} = \int_M \left(\sum_{i=1}^{n} |(D_{\tau_i} X)^{\perp}|^2 - \sum_{i,j=1}^{n} (X \cdot B(\tau_i, \tau_j))^2 \right) d\mathcal{H}^n .$$

9.8 REMARK In case $k = 1$ and M is orientable, with continuous unit normal ν , then $X = \zeta\nu$ for some scalar function ζ with compact support on M , and the above identity has the simple form

$$\frac{d^2}{dt^2} \mathcal{H}^n(M_t)\Big|_{t=0} = \int_M (|\nabla^M \zeta|^2 - \zeta^2 |B|^2)\, d\mathcal{H}^n ,$$

where $|B|^2 = \sum_{i,j=1}^{n} |B(\tau_i, \tau_j)|^2 \equiv \sum_{i,j=1}^{n} |\nu \cdot B(\tau_i, \tau_j)|^2$. This is clear, because $(D_{\tau_i}(\nu\zeta))^{\perp} = \nu\, D_{\tau_i}\zeta$ by virtue of the fact that $D_{\tau_i}\nu\big|_y \in T_y M$ $\forall\, y \in M$.

Proof of Lemma 9.7 First we note that $\int_M \operatorname{div}_M Z\, d\mathcal{H}^n = 0$ by virtue of the fact that M is stationary in U , and second we note that $\operatorname{div}_M X = -X \cdot \underline{\underline{H}} = 0$ by virtue of 7.5' and 9.5(2) and the fact that X is normal to M . The proof is completed by noting that $\tau_i \cdot D_{\tau_j} X \equiv X \cdot B(\tau_i, \tau_j)$ by virtue of 7.3 and the fact that X is normal to M .

§10. CO-AREA FORMULA

As in our discussion of the area formula, we begin by looking at linear maps $\lambda : \mathbb{R}^n \to \mathbb{R}^N$, but here we assume $N < n$. Let us first look at the special case when λ is the orthogonal projection p of $\mathbb{R}^n$ onto $\mathbb{R}^N$. (As before, we identify $\mathbb{R}^N$ with the subspace of $\mathbb{R}^n$ consisting of all points

$(x^1,\ldots,x^n)$ with $x^j = 0$, $j = n-N+1,\ldots,n$.) The orthogonal projection p has the property that, for each $y \in \mathbb{R}^N$, $p^{-1}(y)$ is an $(N-n)$-dimensional affine space; each of these spaces is a translate of the $(N-n)$-dimensional subspace $p^{-1}(0)$. Thus the inverse images $p^{-1}(y)$ decompose all of $\mathbb{R}^n$ into parallel "$(n-N)$-dimensional slices" and by Fubini's Theorem

10.1 $$\int_{\mathbb{R}^N} H^{n-N}(p^{-1}(y) \cap A)\,dy = H^n(A)$$

whenever A is an L^n-measurable subset of $\mathbb{R}^n$.

This formula (which, we emphasize again, is just Fubini's Theorem) is a special case of a more general formula known as the co-area formula. We first derive this in case of an arbitrary linear map $\lambda : \mathbb{R}^n \to \mathbb{R}^N$ with rank $\lambda = N$.

Let $F = \lambda^{-1}(0)$. (Then for each $y \in \mathbb{R}^N$, $\lambda^{-1}(y)$ is an $(n-N)$-dimensional affine space which is a translate of F; the sets $\lambda^{-1}(y)$ thus decompose all of $\mathbb{R}^n$ into parallel $(n-N)$-dimensional slices.)

Take an orthogonal transformation $q = \mathbb{R}^n \to \mathbb{R}^n$ such that $q\,F^{\perp} = \mathbb{R}^N$, $q\,F = \mathbb{R}^{n-N}$. Then λ can be represented in the form $\lambda = \sigma \circ p \circ q$, where p is the orthogonal projection $\mathbb{R}^n$ onto $\mathbb{R}^N$ and σ is a non-singular transformation of $\mathbb{R}^N$. (This is easily checked by considering the action of λ on suitable basis vectors.) By 10.1 above, for any H^n-measurable $A \subset \mathbb{R}^n$,

$$L^n(A) = L^n(q(A)) = \int_{\mathbb{R}^N} H^{n-N}(q(A) \cap p^{-1}(y))\,dL^N(y)$$

$$= \int_{\mathbb{R}^N} H^{n-N}(A \cap q^{-1}(p^{-1}(y)))\,dL^N(y) .$$

making the change of variable $z = \sigma(y)$ $(dy = |\det\sigma|^{-1}dz)$, we thus get

$$|\det\sigma|\,\mathcal{L}^n(A) = \int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap q^{-1}(p^{-1}(\sigma^{-1}(z))))\,d\mathcal{L}^N(z)$$

$$\equiv \int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap \lambda^{-1}(z))\,d\mathcal{L}^N(z) \ .$$

Also, since $q^*q = 1_{\mathbb{R}^n}$ and $pp^* = 1_{\mathbb{R}^N}$, we have $\lambda \circ \lambda^* = \sigma \circ \sigma^* : \mathbb{R}^N \to \mathbb{R}^N$, so that $|\det\sigma| = \sqrt{\det \lambda\circ\lambda^*}$.

Thus finally

10.2 $$\sqrt{\det \lambda\circ\lambda^*}\ \mathcal{L}^n(A) = \int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap \lambda^{-1}(z))\,d\mathcal{L}^N(z) \ .$$

This is the co-area formula for linear maps. (Note that it is trivially valid, with both sides zero, in case rank $\lambda < N$.)

Generally, given a C^1 map $f : M \to \mathbb{R}^N$, where M is an n-dimensional C^1 submanifold of $\mathbb{R}^{n+k}$, we can define

$$J^*f(x) = \sqrt{\det(df_x) \circ (df_x)^*} \ ,$$

where, as usual, $df_x : T_xM \to \mathbb{R}^N$ denotes the induced linear map. Then for any Borel set $A \subset M$

10.3 $$\int_A J^*f \, d\mathcal{H}^n = \int_{\mathbb{R}^N} \mathcal{H}^{n-N}(A \cap f^{-1}(y))\ d\mathcal{L}^N(y) \ .$$

This is the general co-area formula. Its proof uses an approximation argument based on the linear case 10.2. (See [HR1] or [FH1] for the details.)

An important consequence of 10.3 is that if $C = \{x \in M : J^*f(x) = 0\}$, then (by using 10.3 with $A = C$) $\mathcal{H}^{n-N}(C \cap f^{-1}(y)) = 0$ for $\mathcal{L}^N$-a.e. $y \in \mathbb{R}^N$. Since $J^*f(x) \neq 0$ precisely when df_x has rank N, it is clear from the implicit function theorem that $x \in f^{-1}(y) \sim C \Rightarrow \exists$ a neighbourhood V of x

such that $V \cap f^{-1}(y)$ is an $(n-N)$-dimensional C^1 submanifold. In summary we thus have the following important result.

10.4 THEOREM (C^1 Sard-type theorem.)

Suppose $f : M \to \mathbb{R}^N$, $N < n$, *is* C^1. *Then for* L^N-a.e. $y \in f(M)$, $f^{-1}(y)$ *decomposes into an* $(n-N)$-*dimensional* C^1 *submanifold and a closed set of* H^{n-N}-*measure zero. Specifically,*

$$f^{-1}(y) = (f^{-1}(y) \sim C) \cup (f^{-1}(y) \cap C) ,$$

$C = \{x \in M : J^*f(x) = 0\}$ $(\equiv \{x \in M : \operatorname{rank}(df_x) < N\})$, $H^{n-N}(f^{-1}(y) \cap C) = 0$, L^N-a.e. y, *and* $f^{-1}(y) \sim C$ *is an* $(n-N)$-*dimensional* C^1 *submanifold.*

10.5 REMARK If f and M are of class C^{n-N+1}, then *Sard's Theorem* asserts the stronger result that in fact $f^{-1}(y) \cap C = \emptyset$ for L^N-a.e. $y \in \mathbb{R}^N$, so that $f^{-1}(y)$ is an $(n-N)$-dimensional C^{n-N+1} submanifold for L^N-a.e. $y \in \mathbb{R}^N$.

A useful generalization of 10.3 is as follows: If g is a non-negative H^n-measurable function on M, then

10.6 $$\int_M (J^*f) g \, dH^n = \int_{\mathbb{R}^N} \int_{f^{-1}(y)} g \, dH^{n-N} \, dL^N(y) .$$

10.7 REMARKS

(1) Notice that the above formulae enable us to bound the H^{n-N} measure of the "slices" $f^{-1}(y)$ for a good set of y. Specifically if $|f| \le R$ and g is as in 10.6 ($g \equiv 1$ is an important case), then there must be a set $S \subset B_R(0)$ $(\subset \mathbb{R}^N)$, $S = S(g,f,M)$, with $L^N(S) \ge \frac{1}{2} L^N(B_R(0))$ and with

$$\int_{f^{-1}(y)} g \, dH^{n-N} \le \frac{2}{L^N(B_R(0))} \int_M g \, J^*f \, dH^n$$

for each $y \in S$. For otherwise there would be a set $T \subset B_R(0)$ with $\mathcal{L}^N(T) > \frac{1}{2}\mathcal{L}^N(B_R(0))$ and

$$\int_{f^{-1}(y)} g \, d\mathcal{H}^{n-N} \geq \frac{2}{\mathcal{L}^N(B_R(0))} \int_M g \, J^*f \, d\mathcal{H}^n \; , \quad y \in T \; ,$$

so that, integrating over T we obtain a contradiction to 10.6 if $\int_M g \, J^*f \, d\mathcal{H}^n > 0$. On the other hand, if $\int_M g \, J^*f \, d\mathcal{H}^n = 0$ then the required result is a trivial consequence of 10.6.

(2) The above has an important extension to the case when we have $f : \mathbb{R}^{n+k} \to \mathbb{R}^n$ and *sequences* $\{M_j\}$, $\{g_j\}$ satisfying the conditions of M, g above. In this case there is a set $S \subset B_R(0)$ with $\mathcal{L}^N(S) \geq \frac{1}{2}\mathcal{L}^N(B_R(0))$ such that for each $y \in S$ there is a *subsequence* $\{j'\}$ (depending on y) with

$$\int_{M_{j'} \cap f^{-1}(y)} g_{j'} \, d\mathcal{H}^{n-N} \leq \frac{2}{\mathcal{L}^N(B_R(0))} \int_{M_{j'}} g_{j'} \, J^*f \, d\mathcal{H}^n \; .$$

Indeed otherwise there is a set T with $\mathcal{L}^N(T) > \frac{1}{2}\mathcal{L}^N(B_R(0))$ so that for each $y \in T$ there is $\ell(y)$ such that

$$(*) \qquad \int_{M_j \cap f^{-1}(y)} g_j \, d\mathcal{H}^{n-N} > \frac{2}{\mathcal{L}^N(B_R(0))} \int_{M_j} g_j \, J^*f \, d\mathcal{H}^n$$

for each $j > \ell(y)$. But $T = \bigcup_{j=1}^{\infty} T_j$, $T_j = \{y \in T : \ell(y) \leq j\}$, and hence there must exist j so that $\mathcal{L}^N(T_j) > \frac{1}{2}\mathcal{L}^N(B_R(0))$. Then, integrating (*) over $y \in T_j$, we obtain a contradiction to 10.6 as before.