Viscosity Solutions of Second Order Partial Differential Equations

by

Kris A. Kissel

B.S., University of Pittsburgh, 1998

Submitted to the Graduate Faculty of
Arts and Sciences in partial fulfillment
of the requirements for the degree of
Master of Science in Mathematics

University of Pittsburgh

2000

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Kris A. Kissel

It was defended on

April 24, 2000

and approved by

Prof. Juan J. Manfredi (Chairman)

Prof. Frank H. Beatrous

Prof. Thomas A. Metzger

Committee Chairperson(s)

Viscosity Solutions of Second Order Partial Differential Equations

Kris A. Kissel, M.S.

University of Pittsburgh, 2000

A general second order partial differential equation (on $\mathbb{R}^n$) can be written as $F(x, u(x), Du(x), D^2u(x)) = 0$ for some function $F : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times S(n) \to \mathbb{R}$, where $S(n)$ is the set of $n \times n$ symmetric real matrices. Here $u : \mathbb{R}^n \to \mathbb{R}$ is the unknown function for which we wish to solve.

A classical solution of this PDE is a function that is twice differentiable which we can plug into the function F as displayed above along with its derivatives to obtain zero on the right side of the equation.

However, this classical understanding is insufficient. There are many PDE for which one can prove that no C^2 function exists that solves the equation above.

One remedy invented to deal with this difficulty is the theory of distributions. But distributions apply primarily to the linear case and do not lend themselves well to nonlinear F.

We must take a local point of view to deal with nonlinear equations. That is to say, we must work with a different kind of generalized derivative that is defined point-wise, instead of globally as in the case with distributions. These generalized pointwise derivatives are called "jets." Looking at jets allows us to define "viscosity solutions" to certain differential equations.

In this thesis, we study a comparison principle in the class of semi-continuous functions: if u s a subsolution and v is a supersolution on $\mathbf{\Omega}$ in the viscosity sense, and $u \leq v$ on $\partial\mathbf{\Omega}$, then $u \leq v$ in $\mathbf{\Omega}$.

We will use several deep results from real analysis about convex functions in the course of proving these results, so we spend the first chapter studying convex functions in preparation.

Contents

1 Convex Functions 1
1.1 Basic Definitions and Well-Known Facts 1
1.2 Convex Functions are Locally Lipschitz 4
1.3 Second Derivatives in the Sense of Measures 14
1.4 Derivatives of Convex Functions are in BV_{loc} 16
1.5 Aleksandrov's Theorem 20
1.6 Jensen's Lemma 30

2 Viscosity Solutions 36
2.1 "Proper" Functions 37
2.2 Subjets and Superjets 39
2.3 Viscosity Solutions 42
2.4 Closures of Jets 43

3 The Comparison Principle 45
3.1 Hamilton-Jacobi Equations 45
3.2 Sup Convolutions 49
3.3 The Maximum Principle 56
3.4 The Comparison Principle 61

Bibliography 69

List of Symbols

$\mathbb{R}$	the real numbers
$\mathbb{R}^n$	n-dimensional euclidean space
$\mathbf{\Omega}$	an open set in $\mathbb{R}^n$
$B(x,r)$	the open ball of radius r centered at x in $\mathbb{R}^n$
$\overline{B}(x,r)$	the closed ball of radius r centered at x in $\mathbb{R}^n$
∂S	the topological boundary of the set $S \subset \mathbb{R}^n$
$\mathcal{L}^n$	n-dimensional Lebesgue measure
$\alpha(n)$	the Lebesgue volume of the unit ball in $\mathbb{R}^n$
$⨍_X f\ dx$	the average value of f over the set X
Df	the gradient of the function f (this may be a classical or weak derivative, depending on context)
$D^2 f$	the hessian matrix of the function f (classical or weak second derivative)
$J^{2,+}u(x)$	the second-order superjet of u at x (see p. 42)
$J^{2,-}u(x)$	the second-order subjet of u at x (see p. 42)
$J^2u(x)$	the second-order jet of u at x (see p. 42)
$\overline{J}^{2,+}$	the closure of the second-order superjet of u at x (see p. 44)
$\overline{J}^{2,-}$	the closure of the second-order subjet of u at x (see p. 44)
$\overline{J}^2u(x)$	the closure of the second-order jet of u at x (see p. 44)
f_λ	the sup-convolution of the function f (see p. 50)
$C^k(\mathbf{\Omega})$	the space of k-times continuously differentiable functions on $\mathbf{\Omega}$ for $k \geq 1$
$C(\mathbf{\Omega})$	the space of continuous functions on $\mathbf{\Omega}$
$C^k_c(\mathbf{\Omega})$	the space of functions in $C^k(\mathbf{\Omega})$ with compact support
$L^p(\mathbf{\Omega})$	the space of functions f on $\mathbf{\Omega}$ such that $\int_{\mathbf{\Omega}} \|f\|^p\ dx < \infty$
$L^p_{loc}(\mathbf{\Omega})$	the space of functions f on $\mathbf{\Omega}$ such that $\int_K \|f\|^p\ dx < \infty$ for each compact set $K \subset \mathbf{\Omega}$
$USC(\mathbf{\Omega})$	the space of upper-semicontinuous functions on $\mathbf{\Omega}$
$LSC(\mathbf{\Omega})$	the space of lower-semicontinuous functions on $\mathbf{\Omega}$
$W^{k,p}(\mathbf{\Omega})$	the space of k-times weakly-differentiable functions in $L^p(\mathbf{\Omega})$ whose weak derivatives are also in $L^P(\mathbf{\Omega})$
$W^{k,p}_{loc}$	the space of k-times weakly-differentiable functions in $L^p_{loc}(\mathbf{\Omega})$ whose weak derivatives are also in $L^p_{loc}(\mathbf{\Omega})$
$\langle \alpha, \beta \rangle$	the inner product $\sum_{i=1}^n \alpha_i \beta_i$ on $\mathbb{R}^n$
η_ϵ	the standard mollifier defined by $\eta(x) = Ce^{(\frac{1}{\|x\|^2-1})}$ if $\|x\| < 1$ and $\eta = 0$ otherwise; then $\eta_\epsilon(x) = \frac{1}{\epsilon^n}\eta(\frac{x}{\epsilon})$; here C is chosen so that $\int_{\mathbb{R}^n} \eta\ dx = 1$

Chapter 1

Convex Functions

When we prove the Maximum Principle later, we will use a technique whereby we approximate a semicontinuous function by its "sup convolution." The reason for this is that, since sup convolutions have "jets" (generalized pointwise first and second derivatives, which will be defined more precisely in Chapter 2) at the right places, then we can show that the original function also has jets at the right places.

Why do we bother with this transformation at all if we still have to look for the same things? We do so because the sup convolution is always semiconvex, and semiconvexity, like convexity, implies second-order differentiability and therefore an abundance of jets.

Our first goal here is to prove some local estimates for convex functions and their derivatives. Then we move toward Aleksandrov's Theorem, which states that a convex function is twice pointwise-differentiable almost everywhere. The last result in this chapter will be Jensen's Lemma, which studies the local maxima we get from a convex function by perturbing it with a linear term.

1.1 Basic Definitions and Well-Known Facts

Definition 1.1 *Let* $\mathbf{\Omega} \subset \mathbb{R}^n$ *be a convex domain. A function* $\phi : \mathbf{\Omega} \to \mathbb{R}$ *is said to be* convex *if for* $x,\ y \in \mathbf{\Omega}$ *and* $0 < t < 1$, *we have*

$$\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y).$$

A function $\phi : \mathbf{\Omega} \to \mathbb{R}$ *is said to be* semiconvex *with convexity constant* $\lambda \geq 0$ *if* $\phi(x) + \frac{\lambda}{2} \mid x \mid^2$ *is convex.*

Definition 1.2 *A function* $\phi : \mathbf{\Omega} \to \mathbb{R}$ *is said to be* Lipschitz continuous *(or just* Lipschitz*) if there exists a constant* $C \geq 0$ *such that for all* $x,\ y \in \mathbf{\Omega}$*, we have* $| \phi(x) - \phi(y) | \leq C \mid x - y \mid$.

Theorem 1.3 (Rademacher's Theorem) *Locally Lipschitz functions are differentiable almost everywhere with respect to Lebesgue measure.*

We will not prove Rademacher's Theorem here. For a modern proof, the reader should consult [E].

We need the following elementary lemma to gain anything from Theorem 1.5 in the next section.

Lemma 1.4 *Convex functions on* $\mathbf{\Omega}$ *are locally bounded and upper-semicontinuous.*

Proof:

It is an elementary result that convex functions are continuous on lines (see [R]), so the lemma is clearly true in $\mathbb{R}^1$. In particular, convex functions are locally bounded in $\mathbb{R}^1$, and we obtain this result in general by induction on the dimension n as follows.

Suppose the result has been proved for a dimension n. Let K be a parallelepiped in $\mathbb{R}^{n+1}$. The sides of the parallelepiped are connected, compact subsets of n-dimensional hyperplanes. So by the inductive hypothesis, a convex function f is bounded on the surface of the box K. Choose $M \geq 0$ such that $|f| \leq M$ on ∂K. Take any $x = (x_1, ..., x_{n+1})$ in the interior of K. The line $L = \{(x_1, ..., x_n, t); t \in \mathbb{R}\}$ intersects the box K at two points, say y and z. There exists λ with $0 < \lambda < 1$ such that $x = \lambda y + (1 - \lambda)z$. Now by convexity of f, we have

$$f(x) \leq \lambda f(y) + (1 - \lambda)f(z) \leq \lambda M + (1 - \lambda)M = M.$$

So f is bounded above.

Let P be a plane that passes through the interior of K such that the $(n+1)$-th coordinate of points in P are identical. We will say that points in K with a greater $(n + 1)$-th coordinate are "above" P, and that points with a lesser final coordinate are "below" P.

Suppose x is above P. Either y or z is below P, and we may assume that y is it. Let w be the point of intersection of the plane P and the line L. There exists a (new) λ with $0 < \lambda < 1$ such that $w = \lambda x + (1-\lambda)y$. By convexity, we have

$$f(w) \leq \lambda f(x) + (1-\lambda) f(y).$$

That implies

$$\begin{aligned} f(x) &\geq \frac{f(w) + (\lambda - 1) f(y)}{\lambda} \\ &\geq f(w) + (\lambda - 1) f(y) \\ &\geq -M + (\lambda - 1)M \\ &\geq -2M. \end{aligned}$$

A similar argument goes through if x is below the plane P. Since P is a connected subset of an n-dimensional hyperplane, f is bounded on P by our inductive hypothesis.

We have thus shown that convex functions are locally bounded for every dimension n.

Fix $\alpha \in \mathbb{R}$ and $x \in \mathbf{\Omega}$ such that $f(x) < \alpha$. Since convex functions are continuous on lines, there exists $\epsilon > 0$ such that $f(x') < \alpha$ for all $x' \in K_\epsilon$; here, K_ϵ is the set of points x' that share at least one coordinate with x and are a distance less than or equal to ϵ from x.

Let $\hat{K}_\epsilon$ be the convex-hull of K_ϵ:

$$\hat{K}_\epsilon = \{\lambda x' + (1-\lambda)x''; 0 \leq \lambda \leq 1 \text{ and } x', x'' \in K_\epsilon\}.$$

It is easy to see that x is an interior point of $\hat{K}_\epsilon$ and that $f(x') \leq \alpha$ for all $x' \in \hat{K}_\epsilon$. This proves that $f^{-1}([-\infty, \alpha))$ is open. Hence f is upper-semicontinuous.

■

Recall that upper-semicontinuous functions are in fact Lebesgue measurable. Combining this knowledge with the lemma above, we see that a convex function f on $\mathbf{\Omega}$ is in, for example, $L^1_{loc}(\mathbf{\Omega})$. This is the point that we will need in the next section.

1.2 Convex Functions are Locally Lipschitz

Theorem 1.5 *Let* $\mathbf{\Omega}$ *be a convex domain, and* $f : \mathbf{\Omega} \to \mathbb{R}$ *be a convex function. Then* f *is locally Lipschitz on* $\mathbf{\Omega}$*, and there exists a constant* C*, depending only on* n*, such that*

$$(i) \quad \sup_{B(x,\frac{r}{2})} | f | \leq C ⨍_{B(x,r)} | f | \, dy \qquad \textit{and}$$
$$(ii) \quad \text{ess sup}_{B(x,\frac{r}{2})} | Df | \leq \frac{C}{r} ⨍_{B(x,r)} | f | \, dy$$

for each ball $B(x,r) \subset \mathbf{\Omega}$.

Proof:

Step 1: Suppose $f \in C^2(\mathbf{\Omega})$ is convex. Fix $x \in \mathbf{\Omega}$. For each $y \in \mathbf{\Omega}$ and for each $\lambda \in (0,1)$, we have

$$f(x + \lambda(y-x)) \leq f(x) + \lambda(f(y) - f(x)),$$

which can be rearranged to read as

$$\frac{f(x+\lambda(y-x)) - f(x)}{\lambda} \leq f(y) - f(x).$$

Hence
$$\langle Df(x),\ y-x \rangle = \lim_{\lambda \to 0} \frac{f(x+\lambda(y-x)) - f(x)}{\lambda} \leq f(y) - f(x).$$

So for all $x,\ y \in \mathbb{R}^n$, we have that $f(y)$ lies above the tangent hyperplane of f at x:

$$f(y) \geq f(x) + \langle Df(x), y-x \rangle. \tag{1.1}$$

Suppose $z \in B(x, \frac{r}{2})$ and $B(x,r) \subset \mathbf{\Omega}$. The same reasoning that lead to equation (1.1) shows $f(y) \geq f(z) + \langle Df(z), y-z \rangle$. Integrating this inequality over the ball

$B(z,\frac{r}{2})$ gives us

$$\int_{B(z,\frac{r}{2})} f(y)\ dy \geq \int_{B(z,\frac{r}{2})} f(z)\ dy + \int_{B(z,\frac{r}{2})} \langle Df(z), y - z\rangle\ dy.$$

Dividing both sides by the volume of $B(z,\frac{r}{2})$ to obtain averages yields

$$\begin{aligned}
\fint_{B(z,\frac{r}{2})} f(y)\ dy &\geq f(z) + \fint_{B(z,\frac{r}{2})} \langle Df(z), y - z\rangle\ dy \\
&= f(z) + \fint_{B(0,\frac{r}{2})} \langle Df(z), y\rangle\ dy \\
&= f(z).
\end{aligned}$$

This submean-value property for smooth convex functions leads to

$$f(z) \leq \fint_{B(z,\frac{r}{2})} f(y)\ dy \leq 2^n \fint_{B(x,r)} \mid f(y) \mid\ dy. \tag{1.2}$$

Choose a standard cut-off function $\zeta \in C_c^\infty(\mathbb{R}^n)$ such that $0 \leq \zeta \leq 1$, $\mid D\zeta \mid \leq \frac{2^n}{r}$, $\zeta = 1$ on $B(x,\frac{r}{2})$, and $\zeta = 0$ on $\mathbb{R}^n \backslash B(x,r)$. We will use this cut-off function in the normal way it is used to study solutions of elliptic equations: multiply f by the cut-off function and integrate by parts.

By equation 1.1 again, we have $f(z) \geq f(y) + \langle Df(y), z - y\rangle$. Multiply both sides by $\zeta(y)$ and integrate with respect to y over $B(x,r)$ to obtain

$$f(z)\int_{B(x,r)} \zeta(y)\ dy \geq \int_{B(x,r)} f(y)\zeta(y)\ dy + \int_{B(x,r)} \langle Df(y), z - y\rangle \zeta(y)\ dy. \tag{1.3}$$

Now we use integration by parts on the second integral on the right:

$$\begin{aligned}
&\int_{B(x,r)} \langle Df(y), (z - y)\zeta(y)\rangle\ dy \\
&\qquad = -\int_{B(x,r)} f(y)\,\mathrm{div}(\zeta(y)(z - y))\ dy + \int_{\partial B(x,r)} f(y)\zeta(y)\langle z - y, \nu\rangle\ dS(y),
\end{aligned}$$

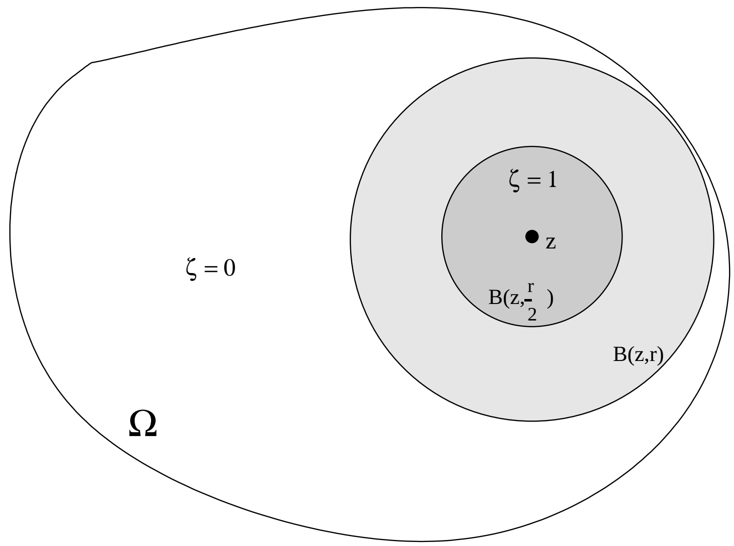


Figure 1.1: The cut-off function ζ is equal to 1 on $B(z, \frac{r}{2})$ and equal to 0 on the domain $\mathbf{\Omega}$ outside $B(z,r)$. It decreases smoothly between these regions.

where ν is the outward-pointing unit normal vector on $\partial B(x,r)$, and where we have used the divergence theorem to obtain the surface integral in the last line. But by our choice of ζ, the integrand is zero on the surface $\partial B(x,r)$, hence the integral is zero. This leaves just the term $-\int_{B(x,r)} f(y)\,\mathrm{div}(\zeta(y)(z-y))\;dy$. Making this substitution in equation 1.3 leads to

$$f(z)\int_{B(x,r)} \zeta(y)\;dy \geq \int_{B(x,r)} f(y)[\zeta(y) - \mathrm{div}(\zeta(y)(z-y))]\;dy. \tag{1.4}$$

We next wish to bound this integral below by a negative constant multiple of

$$\int_{B(x,r)} \mid f(y) \mid \;\; dy.$$

The point is that we will end up with a bound for $f(z)$ from below. Carrying out the differentiation gives us

$$\mathrm{div}(\zeta(y)(z-y)) = -n\zeta(y) + \langle z-y, D\zeta(y)\rangle.$$

Consequently, we have

$$\begin{aligned} |\operatorname{div}(\zeta(y)(z-y)| &\leq n\zeta(y)| + |\langle z-y, D\zeta(y)\rangle| \\ &\leq n + |z-y||D\zeta(y)| \\ &\leq n + |z-y|\frac{2^n}{r} \\ &\leq n + \frac{3r}{2}\frac{2^n}{r} \\ &= n + 3\cdot 2^{n-1}. \end{aligned}$$

This shows us that

$$\zeta(y) - \operatorname{div}(\zeta(y)(z-y)) \leq |\zeta(y)| + |\operatorname{div}(\zeta(y)(z-y))| \leq 1 + n + 3\cdot 2^{n-1},$$

and we can therefore write

$$\begin{aligned} -\int_{B(x,r)} & f(y)[\zeta(y) - \operatorname{div}(\zeta(y)(z-y))]\ dy \\ &\leq \int_{B(x,r)} |f(y)||\zeta(y) - div(\zeta(y)(z-y))|\ \ dy \\ &\leq (1 + n + 3\cdot 2^{n-1})\int_{B(x,r)} |f(y)|\ \ dy. \end{aligned}$$

Divide both ends of this expression by -1 to obtain a lower bound for the integral on the left. Then substitute the result back in to 1.4 to find

$$\begin{aligned} f(z)\int_{B(x,r)} \zeta(y)\ dy &\geq \int_{B(x,r)} f(y)[\zeta(y) - \operatorname{div}(\zeta(y)(z-y))]\ dy \\ &\geq -(1 + n + 3\cdot 2^{n-1})\int_{B(x,r)} |f(y)|\ \ dy. \end{aligned}$$

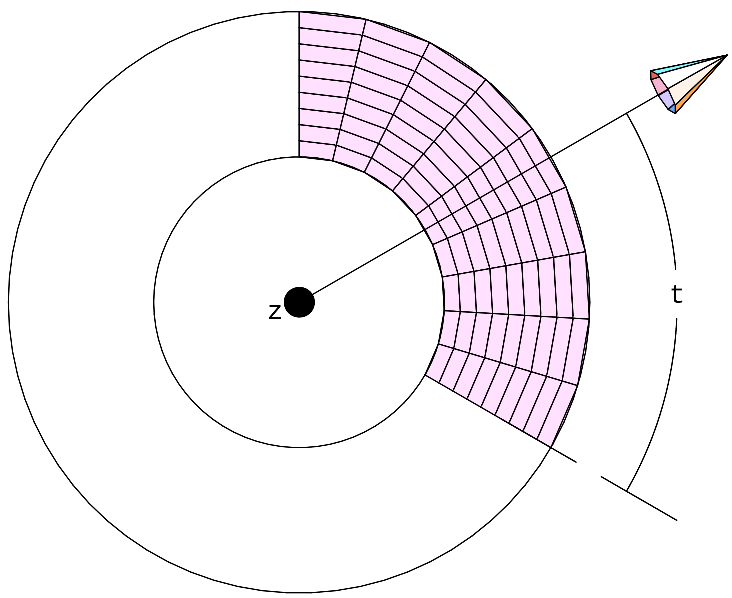


Figure 1.2: The shaded region is S_z, depicted here in the case $n = 2$. It is the sector of the sphere $\overline{B}(z, \frac{r}{2}) \backslash B(z, \frac{r}{4})$ swept out to an angle $t = \frac{\pi}{3}$ from the vector $Df(z)$.

Here we have bounded $f(z)$ below by a constant multiple of $⨍_{B(x,r)} |f(y)|\ dy$, and in 1.2 we did the same for an upper bound. Combining these two results, we have proved

$$| f(z) | \leq C_1 ⨍_{B(x,r)} | f(y) | \ dy. \tag{1.5}$$

This proves (i) for $f \in C^2(\mathbf{\Omega})$.

Step 2: Let $z \in B(x, \frac{r}{2})$. Put

$$S_z = \left\{ y \in \mathbb{R}^n; \frac{r}{4} \leq | y - z | \leq \frac{r}{2} \text{ and } \langle Df(z), y - z \rangle \geq \frac{1}{2} \mid Df(z) \mid\mid y - z \mid \right\}.$$

If $Df(z) = 0$, then S_z is just the hollowed out sphere $\frac{r}{4} \leq | y - z | \leq \frac{r}{2}$. On the other hand, if $Df(z)$ is not 0, then this is still a set of positive volume. Observe that, for any two vectors $\alpha,\ \beta \in \mathbb{R}^n$, we have $\langle \alpha, \beta \rangle = \mid \alpha \mid\mid \beta \mid \cos\theta$, where θ is the angle between α and β. Therefore $\langle \alpha, \beta \rangle \geq \frac{1}{2} \mid \alpha \mid\mid \beta \mid$ whenever $0 \leq \theta \leq \frac{\pi}{3}$. Therefore S_z is a sector of the hollowed out sphere described above, and it is the sector swept out by an angle $\frac{\pi}{3}$, and as such S_z has positive volume. The volume of S_z depends only on n, the dimension of the space. We can now calculate that, for all $y \in S_z$,

$$\begin{aligned} f(y) &\geq f(z) + \langle Df(z), (y-z)\rangle \qquad \text{(since } f \text{ is convex)} \\ &\geq f(z) + \frac{1}{2} \mid Df(z) \mid\mid y - z \mid \qquad \text{(since } y \in S_z) \\ &\geq f(z) + \frac{1}{2} \mid Df(z) \mid \frac{r}{4} \qquad \text{(again, since } y \in S_z) \\ &= f(z) + \frac{r}{8} \mid Df(z) \mid, \end{aligned}$$

and thus

$$\mid Df(z) \mid \leq \frac{8}{r} \left(f(y) - f(z) \right),$$

so by integrating over S_z and noting that $S_z \subset B(x,r)$, we obtain

$$\begin{aligned} \int_{S_z} \mid Df(z) \mid \ dy &\leq \frac{8}{r} \int_{S_z} \mid f(y) - f(z) \mid \ dy \\ &\leq \frac{8}{r} \int_{B(x,r)} \mid f(y) - f(z) \mid \ dy. \end{aligned}$$

Since $Df(z)$ is constant, we get the upper bound

$$\begin{aligned} \mid Df(z) \mid &\leq \frac{8}{\mathcal{L}^n(S_z) r} \int_{B(x,r)} \mid f(y) - f(z) \mid dy \\ &\leq \frac{8}{\mathcal{L}^n(S_z) r} \left(\int_{B(x,r)} \mid f(y) \mid \ dy + \mid f(z) \mid \mathcal{L}^n(B(x,r)) \right) \\ &\leq \frac{8}{\mathcal{L}^n(S_z) r} \left(\int_{B(x,r)} \mid f(y) \mid \ dy + C \int_{B(x,r)} \mid f(y) \mid \ dy \right) \qquad \text{(by 1.5)} \\ &\leq \frac{C}{r} \int_{B(x,r)} \mid f(y) \mid \ dy. \end{aligned}$$

This completes the proof of (ii) for $f \in C^2(\mathbf{\Omega})$.

Step 3: Now we just assume that f is convex, not necessarily smooth. Define $f^\epsilon = \eta_\epsilon * f$, where $\epsilon > 0$ and η_ϵ is the standard mollifier. (The operation $*$ is convolution.) Fix $x,\ y \in \mathbb{R}^n,\ 0 \leq \lambda \leq 1$. For each $z \in \mathbb{R}^n$,

$$\begin{aligned} f(z-(\lambda x+(1-\lambda)y) &= f(\lambda(z-x)+(1-\lambda)(z-y)) \\ &\leq \lambda f(z-x)+(1-\lambda)f(z-y). \end{aligned}$$

Thus,

$$f(z-(\lambda x+(1-\lambda)y)\eta_\epsilon(z) \leq \lambda f(z-x)\eta_\epsilon + (1-\lambda)f(z-y)\eta_\epsilon(z),$$

and integrating this gives us

$$\begin{aligned} &\int_{\mathbb{R}^n} f(z-(\lambda x+(1-\lambda)y))\eta_\epsilon(z)\ dz \\ &\quad \leq \int_{\mathbb{R}^n} \lambda f(z-x)\eta_\epsilon(z)\ dz + \int_{\mathbb{R}^n} (1-\lambda)f(z-y)\eta_\epsilon(z)\ dy. \end{aligned}$$

That is, $f^\epsilon(\lambda x+(1-\lambda)y) \leq \lambda f^\epsilon(x)+(1-\lambda)f^\epsilon(y)$. This says that f^ϵ is convex.

Since f^ϵ is smooth, we can use the estimates proved above:

$$\mid f^\epsilon(z) \mid \leq C_1 ⨍_{B(x,r)} \mid f^\epsilon(y) \mid\ dy$$

and

$$\mid Df^\epsilon(z) \mid \leq \frac{C_2}{r} ⨍_{B(x,r)} \mid f^\epsilon(y) \mid\ dy$$

for $z \in B(x, \frac{r}{2})$. Because $f^\epsilon \to f$ a.e., we have

$$\mid f(z) \mid \leq C_1 ⨍_{B(x,r)} \mid f(y) \mid\ dy$$

for a.e. $z \in B(x, \frac{r}{2})$ since $f \in L^1_{loc}(\mathbf{\Omega})$ by Lemma 1.4, and this implies $f^\epsilon \to f$ in L^1_{loc}.

We claim that in fact $\mid f(z) \mid \leq C_1 ⨍_{B(x,r)} \mid f(y) \mid \ dy$ for all $z \in B(x, \frac{r}{2})$. This will follow from the claim that f is continuous.

We know that $\mid Df^\epsilon \mid \leq \frac{C_2}{r} ⨍_{B(x,r)} \mid f^\epsilon(y) \mid \ dy$, and that

$$⨍_{B(x,r)} \mid f^\epsilon(y) \mid \ dy \to ⨍_{B(x,r)} \mid f(y) \mid \ dy < \infty \quad \text{as } \epsilon \to 0.$$

Put $\epsilon_j = \frac{1}{j}$. Then there exists $A > 0$ such that $⨍_{B(x,r)} \mid f^{\epsilon_j}(y) \mid \ dy < A$ for all $j \in \mathbb{N}$. Thus

$$\mid Df^{\epsilon_j} \mid \leq \frac{C_2 A}{r}, \tag{1.6}$$

for all $j \in \mathbb{N}$. This says the family of functions $\{f^{\epsilon_j}\}_{j \in \mathbb{N}}$ is equicontinuous on every compact set in $B(x, \frac{r}{2})$. Each f^{ϵ_j} is continuous, it is pointwise bounded (by equation 1.6 and the pointwise convergence at some point), so $\{f^{\epsilon_j}\}$ contains a uniformly convergent subsequence. The limit of this sequence must be f. But a uniform limit of continuous functions is continuous. Therefore f is continuous.

This proves (i) for general convex functions.

Step 4: Observe that

$$\sup_{B(x,\frac{r}{2})} \mid Df^\epsilon \mid \leq \frac{C_2}{r} ⨍_{B(x,r)} \mid f^\epsilon(y) \mid \ dy \tag{1.7}$$

implies f^ϵ is Lipschitz. That is, there exists $K(\epsilon) > 0$ such that

$$\mid f^\epsilon(y_1) - f^\epsilon(y_2) \mid \leq K(\epsilon) \mid y_1 - y_2 \mid$$

for all $y_1, \ y_2 \in B(x, \frac{r}{2})$. We can take $K(\epsilon) = \frac{C_2}{r} ⨍_{B(x,r)} \mid f^\epsilon(y) \mid \ dy$, and noting that $⨍_{B(x,r)} \mid f^\epsilon(y) \mid \ dy < A$ as before, we get

$$\mid f^\epsilon(y_1) - f^\epsilon(y_2) \mid \leq \frac{C_2 A}{r} \mid y_1 - y_2 \mid .$$

Using the pointwise convergence of f^ϵ to f as $\epsilon \to 0$, we obtain

$$| f(y_1) - f(y_2) | \leq \frac{C_2 A}{r} | y_1 - y_2 | .$$

So f is locally Lipschitz, and by Rademacher's Theorem, Df exists almost everywhere. By 1.6, whenever Df does exist, $| Df | \leq \frac{C_2 A}{r}$, so $Df \in L^1(B(x, \frac{r}{2}))$, just like f. That is, $f \in W^{1,1}(B(x, \frac{r}{2}))$, hence $f^\epsilon \to f$ in $W^{1,1}_{loc}(B(x, \frac{r}{2}))$, and $Df^\epsilon \to Df$ a.e.in $B(x, \frac{r}{2})$. Taking limits were they exist in 1.7 as $\epsilon \to 0$, we have

$$\text{ess sup}_{B(x,\frac{r}{2})} | Df | \leq \frac{C_2}{r} ⨍_{B(x,r)} | f(y) | \; dy,$$

which completes the proof of (ii) for general convex functions.

■

Theorem 1.6 *Suppose $f \in C^2(\mathbf{\Omega})$ is convex. Then $D^2 f \geq 0$ on $\mathbf{\Omega}$.*

Proof: By Taylor's theorem, we can write

$$\begin{aligned} f(y) = f(x) &+ \langle Df(x), y - x \rangle \\ &+ \left\langle (y-x)^T \cdot \int_0^1 (1-s) D^2 f(x + s(y-x)) \; ds, y - x \right\rangle . \end{aligned}$$

Since $f(y) \geq f(x) + \langle Df(x) y - x \rangle$ for convex f, we then have

$$\begin{aligned} &\left\langle (y-x)^T \cdot \int_0^1 (1-s) D^2 f(x + s(y-x)) \; ds, y - x \right\rangle \\ &= f(y) - f(x) - \langle Df(x), y - x \rangle \\ &\geq 0. \end{aligned}$$

Let $\xi \in \mathbb{R}^n$. Put $y = x + t\xi$ for $t > 0$. Then we have $y - x = x + t\xi - x = t\xi$ and $x + s(y - x) = x + st\xi$. Hence

$$\begin{aligned}
&\left\langle (y-x)^T \cdot \int_0^1 (1-s) D^2 f(x + s(x-y))\ ds, y - x \right\rangle \\
&\quad = \left\langle (t\xi)^T \cdot \int_0^1 (1-s) D^2 f(x + st\xi)\ ds, t\xi \right\rangle \\
&\quad \geq 0.
\end{aligned}$$

Dividing both sides by t, we obtain

$$\left\langle \xi^T \cdot \int_0^1 (1-s) D^2 f(x + st\xi)\ ds, \xi \right\rangle \geq 0.$$

As $t \to 0$, $D^2 f(x + st\xi) \to D^2 f(x)$ uniformly on $s \in [0, 1]$. So taking the limit as $t \to 0$ gives us

$$\begin{aligned}
0 &\leq \left\langle \xi^T \cdot \int_0^1 (1-s) D^2 f(x)\ ds, \xi \right\rangle \\
&= \left\langle \xi^T \cdot D^2 f(x) \cdot \int_0^1 (1-s)\ ds, \xi \right\rangle \\
&= \langle \xi^T \cdot D^2 f(x) \cdot \frac{1}{2}, \xi \rangle \\
&= \frac{1}{2} \langle \xi^T \cdot D^2 f(x), \xi \rangle.
\end{aligned}$$

Therefore $\langle \xi^T \cdot D^2 f(x), \xi \rangle \geq 0$, for all $\xi \in \mathbb{R}^n$.

■

1.3 Second Derivatives in the Sense of Measures

Theorem 1.7 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then there exist signed Radon measures $\mu^{ij} = \mu^{ji}$ such that*

$$\int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_i \partial x_j} \, dx = \int_{\mathbb{R}^n} \phi \, d\mu^{ij} \qquad (i = 1, ..., n)$$

for all $\phi \in C_c^2(\mathbb{R}^n)$. Furthermore, the measures μ^{ii} are nonnegative, for $i = 1, ..., n$.

Proof: Fix $\xi \in \mathbb{R}^n$ with $| \xi |= 1$, and write $\xi = (\xi_1, ..., \xi_n)$. Let η_ϵ be the standard mollifier. Write $f^\epsilon = \eta_\epsilon * f$. Then f^ϵ is smooth and convex, so $D^2 f^\epsilon \geq 0$. Let $\phi \in C_c^2(\mathbb{R}^n)$ with $\phi \geq 0$. Then for each for each $i, j \in \{1, ..., n\}$,

$$\int_{\mathbb{R}^n} f^\epsilon \frac{\partial^2 \phi}{\partial x_i \partial x_j} \xi_i \xi_j \, dx = - \int_{\mathbb{R}^n} \frac{\partial f^\epsilon}{\partial x_j} \frac{\partial \phi}{\partial x_j} \xi_i \xi_j \, dx = \int_{\mathbb{R}^n} \phi \frac{\partial^2 f^\epsilon}{\partial x_i \partial x_j} \xi_i \xi_j \, dx.$$

Summing over i and j, we get

$$\sum_{i,j=1}^{n} \int_{\mathbb{R}^n} f^\epsilon \frac{\partial^2 \phi}{\partial x_i \partial x_j} \xi_i \xi_j \, dx = \int_{\mathbb{R}^n} \phi \langle \xi, D^2 f^\epsilon \xi \rangle \, dx \geq 0. \tag{1.8}$$

Note that the final inequality follows because we chose $\phi \geq 0$ and because $D^2 f^\epsilon \geq 0$.

Now since f is convex, f is locally Lipschitz. In particular, f is continuous. Thus $f^\epsilon \to f$ uniformly on compact subsets of $\mathbb{R}^n$.

Put $L_\xi(\phi) = \sum_{i,j=1}^n \int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_i \partial x_j} \xi_i \xi_j \, dx$. Then by 1.8 and the uniform convergence of $f^\epsilon \to f$ on the support of ϕ, we have

$$L_\xi(\phi) = \lim_{\epsilon \to 0} \sum_{i,j=1}^{n} \int_{R^n} f \frac{\partial^2 \phi}{\partial x_i \partial x_j} \xi_i \xi_j \, dx \geq 0.$$

L_ξ is a positive linear functional on $C_c^\infty(\mathbb{R}^n)$ and it satisfies the hypothesis of the Riesz Representation Theorem. Therefore there exists a Radon measure μ^ξ on $\mathbb{R}^n$ so that

$$L_\xi(\phi) = \int_{\mathbb{R}^n} \phi \, d\mu^\xi \qquad \text{for all } \phi \in C_c^\infty(\mathbb{R}^n).$$

Let $\{e_k\}$ be the standard basis vectors for $\mathbb{R}^n$. Set $\mu^{ii} = \mu^{e_i}$ for $i = 1, ..., n$. For $i \neq j$, set $\xi = \frac{e_i + e_j}{\sqrt{2}}$ to obtain

$$\begin{aligned}
\sum_{k,l=1}^{n} \frac{\partial^2 \phi}{\partial x_k \partial x_l} \xi_k \xi_l &= \sum_{k=1}^{n} \sum_{l \in \{i,j\}} \frac{\partial^2 \phi}{\partial x_k \partial x_l} \xi_k \xi_l \\
&= \sum_{k \in \{i,j\}} \sum_{l \in \{i,j\}} \frac{\partial^2 \phi}{\partial x_k \partial x_l} \xi_k \xi_l \\
&= \sum_{k \in \{i,j\}} \left(\frac{1}{\sqrt{2}} \frac{\partial^2 \phi}{\partial x_k \partial x_i} \xi_k + \frac{1}{\sqrt{2}} \frac{\partial^2 \phi}{\partial x_k \partial x_j} \xi_k \right) \\
&= \frac{1}{2} \frac{\partial^2 \phi}{\partial x_i \partial x_i} + \frac{1}{2} \frac{\partial^2 \phi}{\partial x_i \partial x_j} + \frac{1}{2} \frac{\partial^2 \phi}{\partial x_j \partial x_i} + \frac{1}{2} \frac{\partial^2 \phi}{\partial x_j \partial x_j} \\
&= \frac{1}{2} \left(\frac{\partial^2 \phi}{\partial x_i \partial x_i} + 2 \frac{\partial^2 \phi}{\partial x_i \partial x_j} + \frac{\partial^2 \phi}{\partial x_j \partial x_j} \right).
\end{aligned}$$

With these definitions for the measures μ^{ii} and μ^{ξ} $(= \mu^{\xi(i,j)})$, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_i \partial x_i} \, dx &= \int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_i \partial x_i} (e_i)_i (e_i)_i \, dx \\
&= \sum_{k,l=1}^{n} \int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_k \partial x_l} (e_i)_k (e_i)_j \, dx \\
&= \int_{\mathbb{R}^n} \phi \, d\mu^{ii},
\end{aligned}$$

and for $i \neq j$,

$$\begin{aligned}
\int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_i \partial x_j} \, dx &= \int_{\mathbb{R}^n} f \sum_{k,l=1}^{n} \frac{\partial^2 \phi}{\partial x_k \partial x_l} \xi_k \xi_l \, dx - \frac{1}{2} \left(\int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_i \partial x_i} \, dx + \int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_j \partial x_j} \, dx \right) \\
&= \int_{\mathbb{R}^n} \phi \, d\mu^{\xi} - \frac{1}{2} \int_{\mathbb{R}^n} \phi \, d\mu^{ii} - \frac{1}{2} \int_{\mathbb{R}^n} \phi \, d\mu^{jj}.
\end{aligned}$$

So if we set $\mu^{ij} = \mu^{\xi} - \frac{1}{2}\mu^{ii} - \frac{1}{2}\mu^{jj}$ for $i \neq j$, we can write

$$\int_{\mathbb{R}^n} f \frac{\partial^2 \phi}{\partial x_i \partial x_j} \, dx = \int_{\mathbb{R}^n} \phi \, d\mu^{ij},$$

and this expression holds for all $i, j \in \{i, ..., n\}$.

As a final note, observe that we made the measures μ^{ii} nonnegative when we defined $\mu^{ii} = \mu^{e_i}$.

■

1.4 Derivatives of Convex Functions are in BV_{loc}

Definition 1.8 *A function $f \in L^1_{loc}(\mathbf{\Omega})$ has* locally bounded variation *in $\mathbf{\Omega}$ if for each open set $V \subset\subset \mathbf{\Omega}$,*

$$\sup\left\{\int_V f \ \operatorname{div}(\phi)\ dx;\ \phi \in C^1_c(V;\mathbb{R}^n), \mid \phi \mid \leq 1\right\} < \infty.$$

Write $BV_{loc}(\mathbf{\Omega})$ to denote the space of functions with locally bounded variation on $\mathbf{\Omega}$.

Theorem 1.9 (Structure Theorem for BV_{loc} Functions) *If $f \in BV_{loc}(\mathbf{\Omega})$, then there exist a Radon measure μ on $\mathbf{\Omega}$ and a μ-measurable function $\sigma : \mathbf{\Omega} \to \mathbb{R}^n$ such that*

(i) $\mid \sigma(x) \mid = 1$ *a.e. with respect to μ and*

(ii) $$\int_{\mathbf{\Omega}} f \ \operatorname{div}(\phi)\ dx = -\int_{\mathbf{\Omega}} \phi \cdot \sigma \ d\mu$$

for all $\phi \in C^1_c(\mathbf{\Omega};\mathbb{R}^n)$.

Proof: Define a linear functional $L : C^1_c(\mathbf{\Omega};\mathbb{R}^n) \to \mathbb{R}$ by

$$L(\phi) = -\int_{\mathbf{\Omega}} f \ \operatorname{div}(\phi)\ dx,$$

for $\phi \in C^1_c(\mathbf{\Omega};\mathbb{R}^n)$. L is bounded since

$$C(V) = \sup\left\{L(\phi);\ \phi \in C^1_c(V;\mathbb{R}^n), \mid \phi \mid \leq 1\right\} < \infty$$

for each open set $V \subset\subset \mathbf{\Omega}$. Therefore we can use continuity to extend L uniquely to a linear functional $\overline{L} : C_c(\mathbf{\Omega}; \mathbb{R}^n) \to \mathbb{R}$, for which we have

$$\sup \left\{ \overline{L}(\phi);\ \phi \in C_c(\mathbf{\Omega}; \mathbb{R}^n), \mid \phi \mid \leq 1,\ \text{support}\ \ \phi \subset K \right\} < \infty.$$

This is true for every compact set $K \subset \mathbf{\Omega}$.

Therefore, by the Riesz Representation Theorem, there exist a Radon measure μ on $\mathbf{\Omega}$ and a μ-measurable function $\sigma : \mathbb{R}^m \to \mathbb{R}^n$ such that

$$\mid \sigma(x) \mid = 1 \ a.e.\ [\mu] \ and$$

$$L(\phi) = \int_{\mathbf{\Omega}} \phi \cdot \sigma \ d\mu$$

for all $\phi \in C_c(\mathbf{\Omega}; \mathbb{R}^n)$. This is exactly what we wanted to prove.

■

Notation: If $f \in BV_{loc}(\mathbf{\Omega})$, we will henceforth write $\parallel Df \parallel$ for the measure μ and $[Df] = \sigma \parallel Df \parallel$, so that

$$\int_{\mathbf{\Omega}} f \ \ \text{div}(\phi) \ dx = -\int_{\mathbf{\Omega}} \phi \cdot \sigma \ d \parallel Df \parallel = -\int_{\mathbf{\Omega}} \phi \ d[Df]$$

for all $\phi \in C_c^1(\mathbf{\Omega}; R^n)$.

We also write $\mu^i = \sigma^i \parallel Df \parallel$ for $i = 1, ..., n$. Using Lebesgue's Decomposition Theorem, we set $\mu^i = \mu^i_{ac} + \mu^i_s$, where $\mu^i_{ac} \ll \mathcal{L}^n$ and $\mu^i_s \perp \mathcal{L}^n$. Then

$$\mu^i_{ac} = f_i \mathcal{L}^n$$

for some function $f_i \in L^1_{loc}(\mathbf{\Omega})$.

Further, we denote

$$Df = \left(\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n}\right) = (f_1, ..., f_n)$$

$$[Df]_{ac} = (\mu^1_{ac}, ..., \mu^n_{ac}) = Df\mathcal{L}^n, \qquad \text{and}$$

$$[Df]_s = (\mu^1_s, ..., \mu^n_s)$$

Thus we see that the distributional-derivative of f satisfies

$$[Df] = [Df]_{ac} + [Df]_s = Df\mathcal{L}^n + [Df]_s,$$

so that $Df \in L^1_{loc}(\mathbf{\Omega}; \mathbb{R}^n)$ is the density of the absolutely continuous part of the measure $[Df]$.

Theorem 1.10 *Let $f : \mathbf{\Omega} \to \mathbb{R}$ be convex. Then $\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n} \in BV_{loc}(\mathbf{\Omega})$.*

Proof: Let $V \subset\subset \mathbf{\Omega}$, $\phi \in C^2_c(V; \mathbb{R}^n)$, $|\ \phi\ | \leq 1$. Then for $k = 1, ..., n$,

$$\begin{aligned}
\int_{\mathbf{\Omega}} \frac{\partial f}{\partial x_k} \operatorname{div}(\phi)\ dx &= -\sum_{i=1}^n \int_{\mathbf{\Omega}} f \frac{\partial^2 \phi^i}{\partial x_i \partial x_k}\ dx \\
&= -\sum_{i=1}^n \int_{\mathbf{\Omega}} \phi^i\ d\mu^{ik} \\
&\leq \sum_{i=1}^n \int_{\mathbf{\Omega}} |\ \phi^i\ |\ \ d\ |\ \mu^{ik}\ | \\
&\leq \sum_{i=1}^n \int_V d\ |\ \mu^{ik}\ | \qquad \text{(since } |\phi| \leq 1 \text{ and } \phi \text{ has support in } V) \\
&\leq \sum_{i=1}^n |\ \mu^{ik}\ | (V) \\
&< \infty.
\end{aligned}$$

The final inequality holds because $|\ \mu^{ik}\ | \leq \mu^{\xi} + \frac{1}{2}\mu^{ii} + \frac{1}{2}\mu^{kk}$, and each of μ^{ξ}, μ^{ii} and μ^{kk} is positive and Radon, so each term of the sum is finite.

■

Notation: For a convex function f, we write

$$[D^2 f] = \begin{pmatrix} \mu^{11} & ... & \mu^{1n} \\ \vdots & & \vdots \\ \mu^{n1} & ... & \mu^{nn} \end{pmatrix} = \Sigma \parallel D^2 f \parallel,$$

where Σ is an $n \times m$ matrix-valued function on $\mathbf{\Omega}$ and is $\parallel D^2 f \parallel$-measurable, with $\mid \Sigma \mid = 1$ a.e. with respect to $\parallel D^2 f \parallel$.

We also write

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \mu^{ij} \qquad (i, j = 1, ..., n).$$

Using Lebesgue's Decomposition Theorem, we set

$$\mu^{ij} = \mu^{ij}_{ac} + \mu^{ij}_{s}$$

where $\mu^{ij}_{ac} \ll \mathcal{L}^n$ and $\mu^{ij}_{s} \perp \mathcal{L}^n$, so that $\mu^{ij}_{ac} = f_{ij}\mathcal{L}^n$ for some $f_{ij} \in L^1_{loc}(\mathbf{\Omega})$, and we can write

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = f_{ij}, \qquad (i = 1, ..., n), \qquad\qquad D^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & ... & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & ... & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix},$$

$$[D^2 f]_{ac} = \begin{pmatrix} \mu^{11}_{ac} & ... & \mu^{1n}_{ac} \\ \vdots & & \vdots \\ \mu^{n1}_{ac} & ... & \mu^{nn}_{ac} \end{pmatrix}, \qquad \text{and} \qquad [D^2 f]_{s} = \begin{pmatrix} \mu^{11}_{s} & ... & \mu^{1n}_{s} \\ \vdots & & \vdots \\ \mu^{n1}_{s} & ... & \mu^{nn}_{s} \end{pmatrix}.$$

1.5 Aleksandrov's Theorem

Theorem 1.11 (Aleksandrov's Theorem) *Let $f : \mathbf{\Omega} \to \mathbb{R}$ be a convex function. Then for $\mathcal{L}^n$ a.e. x, we have*

$$\left| f(y) - f(x) - \langle Df(x), y - x\rangle - \frac{1}{2}\langle (y-x), D^2 f(x)(y-x)\rangle \right|$$

$$= o(\mid y - x \mid^2) \text{ as } y \to x.$$

Note that $D^2 f(x)$ is the object defined in the notation section following Theorem 1.10 – it is the weight of the absolutely continuous part of the distributional derivative of f. (Our aim is to show that, on a set of full measure, this absolutely continuous part is all that matters.)

Proof: By Rademacher's Theorem, $Df(x)$ exists for a.e. x. Recall that we have the bound from Theorem 1.5 that $\text{ess sup}_{B(x,\frac{r}{2})} \mid Df \mid \leq \frac{C}{r} ⨍_{B(x,\frac{r}{2})} \mid f(y) \mid \; dy$, so $Df \in L^\infty_{loc}(\mathbf{\Omega})$. Therefore a.e. x is a Lebesgue point for Df. That is,

$$\lim_{r\to 0} ⨍_{B(x,r)} \mid Df(y) - Df(x) \mid \; dy = 0 \text{ for } a.e.\ x. \tag{1.9}$$

The "second derivative" $D^2 f$ exists a.e., where we have defined $\frac{\partial^2 f}{\partial x_i \partial x_j} = f_{ij}$, so that $\mu^{ij}_{ac} = f_{ij}\mathcal{L}^n$, and $f_{ij} \in L^1_{loc}(\mathbf{\Omega})$. So a.e. x is a Lebesgue point for $D^2 f$, i.e.

$$\lim_{r\to 0} ⨍_{B(x,r)} \mid D^2 f(y) - D^2 f(x) \mid \; dy = 0 \text{ for a.e. } x. \tag{1.10}$$

The matrix $[D^2 f]_s$, whose entries are measures and which is the singular part of $[D^2 f]$, is singular to $\mathcal{L}^n$. Since $[D^2 f]_s$ is a Radon measure, it is also a Borel measure, and the balls $B(x,r)$ shrink nicely to x, so (by Theorem 7.13 of [R]),

$$\lim_{r\to 0} \frac{[D^2 f]_s(B(x,r))}{\mathcal{L}^n(B(x,r))} = 0 \text{ for } \mathcal{L}^n \text{ a.e. } x.$$

Hence

$$\lim_{r \to 0} \frac{[D^2 f]_s(B(x,r))}{r^n} = 0 \text{ for } \mathcal{L}^n \text{ a.e. } x. \tag{1.11}$$

Fix an x such that 1.9, 1.10 and 1.11 all hold (a.e. x will suffice). For simplicity of notation, assume $x = 0$. Choose $r > 0$. Let $f^\epsilon = \eta_\epsilon * f$. Fix $y \in B(r)$. By Taylor's Theorem,

$$\begin{aligned} f^\epsilon(y) &= f^\epsilon(0) + \langle Df^\epsilon(0), y\rangle + \int_0^1 \langle (1-s)y, D^2 f^\epsilon(sy) \cdot y\rangle \, ds \\ &= f^\epsilon(0) + \langle Df^\epsilon(0), y\rangle + \frac{1}{2}\langle y, D^2 f(0) \cdot y\rangle \\ &\qquad + \int_0^1 \langle (1-s)y, [D^2 f^\epsilon(sy) - D^2 f(0)] \cdot y\rangle \, ds, \end{aligned}$$

where we have added and subtracted the term $\frac{1}{2}\langle y, D^2 f(0) \cdot y\rangle$ in the last line.

Fix $\phi \in C_c^2(B(r))$ with $| \phi | \leq 1$. Multiply the equation above by ϕ, average over $B(r)$ and apply a change of variables to get

$$\begin{aligned} &⨍_{B(r)} \phi(y) \left(f^\epsilon(y) - f^\epsilon(0) - \langle Df^\epsilon(0), y\rangle - \frac{1}{2}\langle y, D^2 f(0) \cdot y\rangle \right) \, dy \\ &\qquad = \int_0^1 (1-s) \left(⨍_{B(r)} \langle \phi(y) y, [D^2 f^\epsilon(sy) - D^2 f(0)] \cdot y\rangle \, dy \right) \, ds \\ &\qquad = \int_0^1 \frac{1-s}{s^2} \left(⨍_{B(rs)} \langle \phi(\frac{z}{s}) z, [D^2 f^\epsilon(z) - D^2 f(0)] \cdot z\rangle \, dz \right) \, ds \end{aligned} \tag{1.12}$$

Put $g_\epsilon(s) = \int_{B(rs)} \langle \phi(\frac{z}{s} z, D^2 f^\epsilon(z) \cdot z\rangle \, dz$. Then we manipulate sums and integrate by parts to obtain

$$\begin{aligned} g_\epsilon(s) &= \int_{B(rs)} \sum_{i=1}^n \left((\phi(\frac{z}{s}) z_i)(\sum_{j=1}^n \frac{\partial^2 f^\epsilon(z)}{\partial z_i \partial z_j} z_j) \right) \, dz \\ &= \sum_{i,j=1}^n \int_{B(rs)} \phi(\frac{z}{s}) z_i \frac{\partial^2 f^\epsilon(z)}{\partial z_i \partial z_j} z_j \, dz \end{aligned}$$

$$= -\sum_{i,j=1}^{n} \int_{B(rs)} \frac{\partial^2}{\partial z_i \partial z_j} \left(\phi(\frac{z}{s}) z_i z_j \right) f^\epsilon(z) \; dz$$

$$- = \sum_{i,j=1}^{n} \int_{B(rs)} \frac{\partial^2}{\partial z_i \partial z_j} \left(\phi(\frac{z}{s}) z_i z_j \right) f^\epsilon(z) \; dz$$

$$= \int_{B(rs)} f^\epsilon(z) \sum_{i,j=1}^{n} \frac{\partial^2}{\partial z_i \partial z_j} \left(\phi(\frac{z}{s}) z_i z_j \right) \; dz.$$

Since f is convex, f is continuous, and $f^\epsilon \to f$ uniformly on the support of $\phi(\frac{\cdot}{s})$. Hence, as $\epsilon \to 0$,

$$g_\epsilon(s) \to \int_{B(rs)} f(z) \sum_{i,j=1}^{n} \frac{\partial^2}{\partial z_i \partial z_j} \left(\phi(\frac{z}{s}) z_i z_j \right) \; dz$$

$$= \sum_{i,j=1}^{n} \int_{B(rs)} \phi(\frac{z}{s}) z_i z_j \; d\mu^{ij}$$

$$= \sum_{i,j=1}^{n} \int_{B(rs)} \phi(\frac{z}{s}) z_i z_j \; d\mu^{ij}_{ac} + \sum_{i,j=1}^{n} \int_{B(rs)} \phi(\frac{z}{s}) z_i z_j \; d\mu^{ij}_{s}$$

$$= \sum_{i,j=1}^{n} \int_{B(rs)} \phi(\frac{z}{s}) z_i z_j \frac{\partial^2 f}{\partial z_i} \partial z_j \; dz + \sum_{i,j=1}^{n} \int_{B(rs)} \phi(\frac{z}{s}) z_i z_j \; d\mu^{ij}_{s}$$

$$= \int_{B(rs)} \phi(\frac{z}{s}) \langle z, D^2 f(z) \cdot z \rangle \; dz + \sum_{i,j=1}^{n} \int_{B(rs)} \phi(\frac{z}{s}) z_i z_j \; d\mu^{ij}_{s}.$$

Take absolute values, use the Cauchy-Schwarz inequality and replace $|z|$ by rs to get, for $s > 0$,

$$\frac{\mid g_\epsilon(s) \mid}{s^{n+2}} = \frac{1}{s^{n+2}} \mid \int_{B(rs)} \langle \phi(\frac{z}{s}) z, D^2 f^\epsilon(z) \cdot z \rangle \; dz \mid$$

$$\leq \frac{(rs)^2}{s^{n+2}} \int_{B(rs)} \mid \phi(\frac{z}{s}) \mid \| D^2 f^\epsilon(z) \| \; dz.$$

Use $|\phi| \leq 1$ and the definition of f^ϵ to see

$$\frac{\mid g_\epsilon(s) \mid}{s^{n+2}} \leq \frac{(rs)^2}{s^{n+2}} \int_{B(rs)} \mid D^2 f^\epsilon(z) \mid \; dz$$

$$= \frac{r^2}{s^n} \int_{B(rs)} \left\{ \sum_{i,j=1}^{n} \int_{\mathbf{\Omega}} f(y) \left(\frac{\partial^2}{\partial z_i \partial z_j} \eta_\epsilon(z-y)\ dy \right)^2 \right\}^{\frac{1}{2}}\ dz.$$

Then by the definition of $D^2 f$ as a distributional derivative,

$$\begin{aligned} \frac{|\ g_\epsilon(s)\ |}{s^{n+2}} &\leq \frac{r^2}{s^n} \int_{B(rs)} \left| \int_{\mathbf{\Omega}} \eta_\epsilon(z-y)\ d[D^2 f] \right|\ dz \\ &\leq \frac{r^2}{s^n} \int_{B(rs)} \int_{\mathbf{\Omega}} \eta_\epsilon(z-y)\ d \parallel D^2 f \parallel\ dz. \end{aligned}$$

That η_ϵ is supported in $B(0,\epsilon)$ and $|\eta_\epsilon| \leq 1$ leads to

$$\begin{aligned} \frac{|\ g_\epsilon(s)\ |}{s^{n+2}} &\leq \frac{r^2}{s^n} \int_{\mathbf{\Omega}} \left(\int_{B(rs)\bigcap B(y,\epsilon)} \eta_\epsilon(z-y)\ dz \right)\ d \parallel D^2 f \parallel \\ &\leq \frac{r^2 \parallel \eta \parallel_\infty}{s^n \epsilon^n} \int_{B(rs+\epsilon)} \left(\int_{B(rs)\bigcap B(y,\epsilon)} dz \right)\ d \parallel D^2 f \parallel . \end{aligned}$$

Then by an Euclidean calculation of volumes, we see

$$\frac{|\ g_\epsilon(s)\ |}{s^{n+2}} \leq \frac{r^2 \alpha(n) \parallel \eta \parallel_\infty}{s^n \epsilon^n} \min\{(rs)^n, \epsilon^n\} \parallel D^2 f \parallel (B(rs+\epsilon)).$$

We now claim the next inequality,

$$\frac{|\ g_\epsilon(s)\ |}{s^{n+2}} \leq \frac{r^2 \alpha(n) \parallel \eta \parallel_\infty}{s^n \epsilon^n} \min\{(rs)^n, \epsilon^n\} (\alpha(n)(2+ \mid D^2 f(0) \mid))(rs+\epsilon)^n.$$

To see this, first observe that we assumed 0 to be a Lebesgue point so that

$$\lim_{(rs+\epsilon)\to 0} \frac{|\ [D^2 f]_s\ |\ (B(rs+\epsilon))}{(rs+\epsilon)^n} = 0.$$

Hence there exists $R > 0$ such that for $0, rs+\epsilon < R$,

$$\frac{|\ [D^2 f]_s\ |\ (B(rs+\epsilon))}{(rs+\epsilon)^n} \leq \alpha(n),$$

and then for $0 < rs + \epsilon < R$,

$$| [D^2 f]_s | (B(rs + \epsilon)) \leq \alpha(n)(rs + \epsilon)^n.$$

Also,

$$\begin{aligned}
\lim_{r\to 0} ⨍_{B(rs+\epsilon)} & | D^2 f(y) - D^2 f(0) | \ dy \\
&= \lim_{r\to 0} \frac{1}{\alpha(n)(rs+\epsilon)^n} \int_{B(rs+\epsilon)} | D^2 f(y) - D^2 f(0) | \ dy \\
&= 0.
\end{aligned}$$

Consequently, for $0 < rs + \epsilon < R$, making R smaller now, if necessary, we have

$$\frac{1}{\alpha(n)(rs+\epsilon)^n} \int_{B(rs+\epsilon)} | D^2 f(y) - D^2 f(0) | \ dy \leq 1,$$

that is,

$$\int_{B(rs+\epsilon)} | D^2 f(y) - D^2 f(0) | \ dy \leq \alpha(n)(rs + \epsilon)^n.$$

Therefore we can compute, for $0 < rs + \epsilon < R$,

$$\begin{aligned}
\| D^2 f \| (B(rs + \epsilon)) &= | [D^2 f] | (B(rs + \epsilon)) \\
&\leq | [D^2 f]_{ac} | (B(rs + \epsilon)) + | [D^2 f]_s | (B(rs + \epsilon)) \\
&= \int_{B(rs+\epsilon)} | D^2 f(y) | \ dy + | [D^2 f]_s | (B(rs + \epsilon)) \\
&= \int_{B(rs+\epsilon)} | D^2 f(y) - D^2 f(0) + D^2 f(0) | \ dy \\
&\quad + | [D^2 f]_s | (B(rs + \epsilon))
\end{aligned}$$

$$\leq \alpha(n)(rs+\epsilon)^n$$
$$+ \int_{B(rs+\epsilon)} \mid D^2 f(0) \mid \; dy + \alpha(n)(rs+\epsilon)^n$$
$$= \alpha(n)(rs+\epsilon) + \alpha(n)(rs+\epsilon)^n \mid D^2 f(0) \mid \alpha(n)(rs+\epsilon)^n$$
$$= \alpha(n)(rs+\epsilon)^n \mid D^2 f(0) \mid +2\alpha(n)(rs+\epsilon)^n$$
$$= \alpha(n)(2+ \mid D^2 f(0) \mid)(rs+\epsilon)^n.$$

We now have

$$\begin{aligned}
\frac{\mid g_\epsilon(s) \mid}{s^2} &\leq \frac{r^2\alpha(n) \parallel \eta \parallel_\infty}{s^n \epsilon^n} \mid \{(rs)^n, \epsilon^n\}(\alpha(n)(2+ \mid D^2 f(0) \mid)(rs+\epsilon)^n \\
&= \frac{C_1}{s^n\epsilon^n} \min\{(rs)^n, \epsilon^n\}(rs+\epsilon)^n \\
&\leq C_1 2^n r^n \\
&= C_2,
\end{aligned}$$

with $C_1 = r^2(\alpha(n))^2(2+ \mid D^2 f(0) \mid) \parallel \eta \parallel_\infty)$, and since, if $rs \leq \epsilon$, we have

$$\begin{aligned}
C_1 \frac{\min\{(rs)^n, \epsilon^n\}(rs+\epsilon)^n}{s^n\epsilon^n} &= C_1 \frac{(rs)^n(rs+\epsilon)^n}{s^n\epsilon^n} \\
&\leq C_1 \frac{(rs)^n(2\epsilon)^n}{s^n\epsilon^n} \\
&= C_1 2^n r^n,
\end{aligned}$$

and otherwise, if $\epsilon < rs$, we have

$$C_1 \frac{\min\{(rs)^n, \epsilon^n\}(rs+\epsilon)^n}{s^n\epsilon^n} = C_1 \frac{\epsilon^n(rs+\epsilon)^n}{s^n\epsilon^n} \leq C_1 \frac{\epsilon^n(2rs)^n}{s^n\epsilon^n} = C_1 2^n r^n.$$

So $| g_\epsilon(s) | \leq C_2 s^{n+2}$. Since $s^{n+2} \in L^1(B(rs))$, we may apply Lebesgue's Dominated Convergence Theorem to g_ϵ and let $\epsilon \to 0$ in 1.12 to get

$$
\begin{aligned}
&⨍_{B(r)} \phi(y) \left(f(y) - f(0) - Df(0)\cdot y - \frac{1}{2}\langle y, D^2 f(0)\cdot y\rangle \right) \, dy \\
&\quad = \lim_{\epsilon\to 0} ⨍_{B(r)} \phi(y) \left(f^\epsilon(y) - f^\epsilon(0) - Df^\epsilon(0)\cdot y - \frac{1}{2}\langle y, D^2 f(0)\cdot y\rangle \right) \, dy \\
&\quad = \lim_{\epsilon\to 0} \int_0^1 \frac{1-s}{s^2} \left(⨍_{B(rs)} \phi(\frac{z}{s}) \langle z, [D^2 f^\epsilon(z) - D^2 f^\epsilon(0)]\cdot z\rangle] \, dz \right) \, ds \\
&\quad = \lim_{\epsilon\to 0} \int_0^1 \frac{1-s}{s^2} \left(⨍_{B(rs)} \phi(\frac{z}{s}) \langle z, D^2 f^\epsilon(z)\cdot z\rangle \, dz \right) \, ds \\
&\qquad\quad - \int_0^1 \frac{1-s}{s^2} \left(⨍_{B(rs)} \phi(\frac{z}{s}) \langle z, D^2 f(0)\cdot z\rangle \, dz \right) \, ds \\
&\quad = \int_0^1 \left(\frac{1}{\alpha(n)(rs)^n} \left(\int_{B(rs)} \phi(\frac{z}{s}) \langle z, D^2 f(z)\cdot z\rangle \, dz + \sum_{i,j=1}^{n} \int_{B(rs)} \phi(\frac{z}{s}) z_i z_j \, d\mu_s^{ij} \right) \right) \, ds \\
&\qquad\quad - \int_0^1 \frac{1-s}{s^2} (⨍_{B(rs)} \phi(\frac{z}{s}) \langle z, D^2 f(0)\cdot z\rangle \, dz) \, ds \\
&\quad = \int_0^1 \frac{1-s}{s^2} \left(⨍_{B(rs)} \phi(\frac{z}{s}) \langle z, [D^2 f(z) - D^2 f(0)]\cdot z\rangle \, dz \right) \, ds \\
&\qquad\quad + \int_0^1 \frac{1-s}{\alpha(n) s^{2+n} r^n} \left(\sum_{i,j+1}^{n} \int_{B(rs)} \phi(\frac{z}{s}) z_i z_j \, d\mu_s^{ij} \right) \, ds.
\end{aligned}
$$

We now analyze these integrals to show that each is $o(r^2)$ as $r \to 0$.

$$
\begin{aligned}
&\left| \int_0^1 \frac{1-s}{s^2} \left(⨍_{B(rs)} \phi(\frac{z}{s}) \langle z, [D^2 f(z) - D^2 f(0)]\cdot z\rangle \, dz \right) \, ds \right| \\
&\quad \leq \int_0^1 \frac{1-s}{s^2} \left(\int_{B(rs)} | \phi(\frac{z}{s}) || z || [D^2 f(z) - D^2 f(0)] || z | \, dz \right) \, ds \\
&\quad \leq \int_0^1 \frac{1-s}{s^2} (rs)^2 \left(⨍_{B(rs)} | D^2 f(z) - D^2 f(0) | \, dz \right) \, ds \\
&\quad = r^2 \int_0^1 (1-s) \left(⨍_{B(rs)} | D^2 f(z) - D^2 f(0) | \, dz \right) \, ds.
\end{aligned}
$$

Since the point 0 is a Lebesgue point, $⨍_{B(r)} | D^2 f(z) - D^2 f(0) | \, dz \to 0$ as $r \to 0$. Hence $(1-s)⨍_{B(rs)} | D^2 f(z) - D^2 f(0) | \, dz \to 0$ uniformly as $r \to 0$, for $s \in [0,1]$.

Therefore,

$$\int_0^1 (1-s)\left(⨍_{B(rs)} \mid D^2f(z) - D^2f(0) \mid \ dz\right)\ ds \to 0 \qquad \text{as } r \to 0.$$

Therefore,

$$r^2\int_0^1 (1-s)\left(⨍_{B(rs)} \mid D^2f(z) - D^2f(0) \mid \ dz\right)\ ds = o(r^2) \qquad \text{as } r \to 0.$$

Similarly, we have

$$\begin{aligned}
&\left|\int_0^1 \frac{1-s}{\alpha(n)s^{2+n}r^n}\left(\sum_{i,j=1}^n \int_{B(rs)} \phi(\frac{z}{s}) z_i z_j \ d\mu_s^{ij}\right)\ ds\right| \\
&\quad \leq \int_0^1 \frac{1-s}{\alpha(n)s^{2+n}r^n}\left(\sum_{i,j=1}^n \int_{B(rs)} \mid \phi(\frac{z}{s}) \mid\mid z_i \mid\mid z_j \mid \ d \mid \mu_s^{ij} \mid\right)\ ds \\
&\quad \leq \int_0^1 \frac{1-s}{\alpha(n)s^{2+n}r^n}(rs)^2\left(\sum_{i,j=1}^n \int_{B(rs)} d \mid \mu_s^{ij} \mid\right)\ ds \\
&\quad \leq \frac{r^2n^2}{\alpha(n)}\int_0^1 \frac{1-s}{s^n r^n} \mid [D^2f]_s \mid (B(rs))\ ds \\
&\quad = \frac{r^2n^2}{\alpha(n)}\int_0^1 (1-s)\left(\frac{\mid [D^2f]_s \mid (B(rs))}{(rs)^n}\right)\ ds.
\end{aligned}$$

By assumption, $\frac{|[D^2f]_s|(B(r))}{r^n} \to 0$ as $r \to 0$, so $(1-s)\left(\frac{|[D^2f]_s|(B(rs))}{(rs)^n}\right) \to 0$ uniformly as $r \to 0$, on $s \in [0,1]$. Hence,

$$\int_0^1 (1-s)\left(\frac{\mid [D^2f]_s \mid (B(rs))}{(rs)^n}\right)\ ds \to 0 \qquad \text{as } r \to 0.$$

Therefore,

$$\frac{r^2n^2}{\alpha(n)}\int_0^1 (1-s)\left(\frac{\mid [D^2f]_s \mid (B(rs))}{(rs)^n}\right)\ ds = o(r^2) \qquad \text{as } r \to 0.$$

Take the supremum over all $\phi \in C_c^2(B(r))$ with $\mid \phi \mid \leq 1$ to obtain

$$\fint_{B(r)} \left| f(y) - f(0) - \langle Df(0), y \rangle - \frac{1}{2}\langle y, D^2 f(0) \cdot y \rangle \right| \, dy = o(r^2) \qquad \text{as } r \to 0.$$

Set $h(y) = f(y) - f(0) - \langle Df(0), y\rangle - \frac{1}{2}\langle y, D^2 f(0) \cdot y\rangle$. Then

$$\fint_{B(r)} \mid h(y) \mid \, dy = o(r^2) \qquad as \ r \to 0.$$

Put $g(y) = h(y) + \frac{|D^2 f(0)|}{2} \mid y \mid^2$. Notice that g is convex since $f(y)$ is convex, $f(0)$ is constant, $\langle Df(0), y\rangle$ is linear and $-\frac{1}{2}\langle y, D^2 f(0) \cdot y\rangle + \frac{|D^2 f(0)|}{2} \mid y \mid^2$ is convex, because

$$\frac{\mid D^2 f(0) \mid}{2} I - \frac{D^2 f(0)}{2}$$

is positive semidefinite, as it follows easily from Cauchy-Schwarz.

Since g is convex, there is a constant C such that

$$\begin{aligned}
\text{ess sup}_{B(\frac{r}{2})} \mid Dg \mid &\leq \frac{C}{r} \fint_{B(r)} \mid g \mid \, dy \\
&\leq \frac{C}{r} \fint_{B(r)} \left| h + \frac{\mid D^2 f(0) \mid}{2} \mid y \mid^2 \right| \, dy \\
&\leq \frac{C}{r} \fint_{B(r)} \mid h \mid \, dy + \frac{C}{r} \fint_{B(r)} \frac{\mid D^2 f(0) \mid}{2} \mid y \mid^2 \, dy \\
&\leq \frac{C}{r} \fint_{B(r)} \mid h \mid \, dy + \frac{C \mid D^2 f(0) \mid}{2} r.
\end{aligned}$$

Using the triangle inequality on $\mid Dg \mid = \mid Dh + \mid D^2 f(0) \mid y \mid$ gives us

$$\text{ess sup}_{B(\frac{r}{2})} \mid Dh \mid \leq \frac{C}{r} \fint_{B(r)} \mid h \mid \, dy + \frac{C \mid D^2 f(0) \mid}{2} r + \mid D^2 f(0) \mid \frac{r}{2}.$$

Because Dh can be redefined on a null set, we might as well say

$$\sup_{B(\frac{r}{2})} | Dh | \leq \frac{C}{r} \fint_{B(r)} | h | \; dy + Cr$$

for some new constant C. Fix $0 < \epsilon, \eta < 1$, such that $\eta^{\frac{1}{n}} \leq \frac{1}{2}$. Observe that

$$\begin{aligned} \epsilon r^2 \mathcal{L}^n \{z \in B(r) \,||\, h(z) \,|\geq \epsilon r^2\} &= \int_{\{z\in B(r)||h(z)|\geq \epsilon r^2\}} \epsilon r^2 \; dz \\ &\leq \int_{\{z\in B(r)||h(z)|\geq \epsilon r^2\}} | h | \; dz \\ &\leq \int_{B(r)} | h | \; dz \\ &= o(r^{2+n}) \qquad \text{as } r \to 0. \end{aligned}$$

We now know that $\mathcal{L}^n\{z \in b(r) \,||\, h(z) \,|\geq \epsilon r^2\} = o(r^n)$ as $r \to 0$. Thus for some $r_0 = r_0(\epsilon, \eta)$, we have

$$\mathcal{L}^n\{z \in B(r) \,||\, h(z) \,|\geq \epsilon r^2\} < \eta \mathcal{L}^n(B(r)) \qquad \text{for } 0 < r < r_0.$$

Hence for each $y \in B(\frac{r}{2})$, there exists $z \in B(y, \eta^{\frac{1}{n}} r) \subset B(r)$ such that $| h(z) | \leq \epsilon r^2$. For if not,

$$\mathcal{L}^n\{z \in B(r) \,||\, h(z) \,|\geq \epsilon r^2\} \geq \mathcal{L}^n(B(y, \eta^{\frac{1}{n}} r)) = \alpha(n)(\eta^{\frac{1}{n}} r)^n = \eta \mathcal{L}^n(B(r)).$$

This would be a contradiction; consequently we get the desired z, and we have

$$\begin{aligned} | h(y) | &\leq | h(z) | + | h(y) - h(z) | \\ &\leq \epsilon r^2 + (\eta^{\frac{1}{n}} r) \sup_{B(r)} | Dh | \\ &\leq \epsilon r^2 + (\eta^{\frac{1}{n}} r)(\frac{C}{r} \fint_{B(r)} | h | \; dy + Cr) \end{aligned}$$

$$\leq \epsilon r^2 + C\eta^{\frac{1}{n}} r^2 \qquad \text{(for some new constant C)}$$

$$\leq 2\epsilon r^2,$$

provided we choose η such that $C\eta^{\frac{1}{n}} \leq \epsilon$ and keep r small enough.

We have thus shown that $\sup_{B(\frac{r}{2})} \mid h \mid = o(r^2)$ as $r \to 0$, i.e.,

$$\sup_{B(\frac{r}{2})} \left| f(y) - f(0) - Df(0) \cdot y - \frac{1}{2}\langle y, D^2 f(0) \cdot y\rangle \right| = o(r^2) \qquad \text{as } r \to 0.$$

This result actually holds for every $x \in \mathbf{\Omega}$ that satisfies 1.9, 1.10 and 1.11. Therefore, for $\mathcal{L}^n$ a.e. x, we have

$$\left| f(y) - f(x) - \langle Df(x), y - x\rangle - \frac{1}{2}\langle (y-x), D^2 f(x)(y-x)\rangle \right|$$

$$= o(\mid y - x \mid^2) \text{ as } y \to x.$$

■

1.6 Jensen's Lemma

To finish this chapter, we prove Jensen's Lemma. If we perturb a semi-convex function with a strict local maximum just a little, by adding a linear term to it, we can end up with functions with local maxima near where the original maximum was. In a given ball around the original strict maximum, how many points of the ball can be a local maximum for some linear perturbation? Jensen's Lemma says the set of potential local maxima has positive measure, in a given ball of any radius. This is a key result.

Theorem 1.12 (Jensen's Lemma) *Let $\phi : \mathbf{\Omega} \to \mathbb{R}$ be semiconvex and $\hat{x}$ be a strict local maximum point of ϕ. For $p \in \mathbb{R}^n$, set $\phi_p(x) = \phi(x) + \langle p, x\rangle$. Then for $r,\ \delta > 0$,*

$\overline{B}_\delta = \overline{B}(o, \delta)$, *the set*

$$K = \{x \in \overline{B}(\hat{x}, r); \textit{there exists } p \in \overline{B}_\delta \textit{ for which } \phi_p \textit{ has a local maximum at } x\}$$

has positive measure.

Proof:

Step 1: Assume $\phi \in C^2(\mathbf{\Omega})$ and r is so small that ϕ has $\hat{x}$ as a unique maximum point in $\overline{B}(\hat{x}, r)$. Then there exists $\epsilon > 0$ such that

$$\phi(x) > \sup\{\phi(y); |y - \hat{x}| = r\} + \epsilon,$$

because ϕ, being continuous, must attain a supremum on the circle, and that must be strictly less than the unique maximum $\phi(\hat{x})$.

Choose $\delta > 0$ so that

$$\delta \leq \frac{\epsilon}{3 \sup\{| y |; | y - \hat{x} |= r\}}.$$

Then for $p \in \overline{B}_\delta$ and $x \in \partial B(\hat{x}, r)$, we have

$$\begin{aligned} \phi_p(x) &= \phi(x) + \langle p, x \rangle \\ &\leq \sup\{\phi(x); x \in \partial B(\hat{x}, r)\} + | p || x | \\ &\leq \sup\{\phi(x); x \in \partial B(\hat{x}, r)\} + \frac{\epsilon}{3}. \end{aligned}$$

On the other hand, if we also require that

$$\delta \leq \begin{cases} 1 & if \quad | \hat{x} |= 0 \\ \frac{\epsilon}{2|\hat{x}|} & if \quad | \hat{x} | \neq 0 \end{cases}$$

then we get

$$\phi_p(\hat{x}) = \phi(\hat{x}) + \langle p, \hat{x} \rangle$$

$$> \epsilon + \sup\{\phi(x); x \in \partial B(\hat{x}, r)\} - \mid p \mid\mid \hat{x} \mid$$

$$> \frac{\epsilon}{2} + \sup\{\phi(x); x \in \partial B(\hat{x}, r)\}.$$

Since this says that there is a point in the interior $B(\hat{x}, r)$ that takes a value greater than that achieved by ϕ_p on the boundary $\partial B(\hat{x}, r)$, we have thus shown that every maximum of ϕ_p with respect to $\overline{B}(\hat{x}, r)$ lies in its interior. Now let $x \in B(\hat{x}, r)$ be an interior maximum of ϕ_p for $p \in \overline{B}_\delta$. Then $D\phi_p(x) = 0$, so $D\phi(x) = -p$. Thus $-p \in D\phi(K)$, since $x \in K$. Since all ϕ_p have interior maxima, $-p \in D\phi(K)$ for all $p \in \overline{B}_\delta$. Therefore $\overline{B}_\delta \subset D\phi(K)$.

Let $\lambda > 0$ so that $\phi(x) + \frac{\lambda}{2} \mid x \mid^2$ be convex. (Observe that we cannot take $\lambda = 0$ here, since that would make ϕ convex, and convex functions cannot have strict local maxima.) Then $D^2\phi(x) + \lambda I \geq 0$, by Theorem 1.6. Hence $-\lambda I \leq D^2\phi(x)$.

Since $D^2\phi(x) \leq 0$ at any x which is a local max for some ϕ_p, we have $D^2\phi(x) \leq 0$ for all $x \in K$. Consequently, $-\lambda I \leq D^2\phi(x) \leq 0$ for all $x \in K$.

Fix $x \in K$ and let y_i be an eigenvector for $D^2\phi(x)$ and λ_i be an associated eigenvalue. Then

$$-\lambda \mid y_i \mid^2 \leq \lambda_i \mid y_i \mid^2 \leq 0.$$

Hence $-\lambda \leq \lambda_i \leq 0$, since eigenvectors must be nonzero. This implies $0 \leq\mid \lambda_i \mid\leq\mid \lambda \mid$. Consequently,

$$|\det\ D^2\phi(x)| = |\lambda_1\lambda_2...\lambda_n| \leq |\lambda|^n,$$

where $\lambda_1, ...\lambda_n$ are the n eigenvalues of $D^2\phi(x)$. Now using the change-of-variables formula, we have

$$\mathcal{L}^n(B_\delta) \leq \mathcal{L}^n(D\phi(K)) \leq \int_K \mid \det\ D^2\phi(x) \mid\ dx \leq \int_K \mid \lambda \mid^n\ dx = \mid \lambda \mid^n \mathcal{L}^n(K).$$

Divide the far sides of this equation by $| \lambda |^n$ to get

$$\mathcal{L}^n(K) \geq \frac{\mathcal{L}^n(B_\delta)}{| \lambda |^n}.$$

This shows that $\mathcal{L}^n(K) > 0$ if ϕ is C^2.

Step 2: Now we turn to a general semiconvex function ϕ. Define $\phi_\epsilon = \eta_\epsilon * \phi$, where η_ϵ is the standard mollifier. If $\phi(x)+\frac{\lambda}{2} | x |^2$ is convex, then so is $(\phi(x)+\frac{\lambda}{2} | x |^2)_\epsilon$ (see the proof of Aleksandrov's Theorem). We want to show that is ϕ is semiconvex with convexity constant λ, then so is ϕ_ϵ.

$$\begin{aligned}
\left(\phi(x) + \frac{\lambda}{2} | x |^2\right)_\epsilon &= \int_{B(0,\epsilon)} \eta_\epsilon(y)[\phi(x-y) + \frac{\lambda}{2} | x - y |^2] \; dy \\
&= \int_{B(0,\epsilon)} \eta_\epsilon(y)\phi(x-y) \; dy + \frac{\lambda}{2} \int_{B(0,\epsilon)} \eta_\epsilon(y) \; | x |^2 \; dy \\
&+ \lambda \int_{B(0,\epsilon)} \eta_\epsilon(y)\langle x, y\rangle \; dy + \frac{\lambda}{2} \int_{B(0,\epsilon)} \eta_\epsilon(y) \; | y |^2 \; dy \\
&= \phi_\epsilon(x) + \frac{\lambda}{2} | x |^2 + \lambda \int_{B(0,\epsilon)} \eta_\epsilon(y)\langle x, y\rangle \; dy \\
&= \frac{\lambda}{2} \int_{B(0,\epsilon)} \eta_\epsilon(y) \; | y |^2 \; dy. \qquad (1.13)
\end{aligned}$$

The second to last term vanishes:

$$\int_{B(o,\epsilon)} \langle x, y\rangle \; dy = \int_{B(0,\epsilon)^+} \eta_\epsilon(y)\langle x, y\rangle \; dy + \int_{B(0,\epsilon)^-} \eta_\epsilon(y)\langle x, y\rangle \; dy$$

(where $B(0,\epsilon)^+ = \{x = (x_1, ..., x_n) \in B(0,\epsilon); x_1 \geq 0\}$, and $B(0,\epsilon)^- = B(0,\epsilon)\backslash B(0,\epsilon)^+$)

$$\begin{aligned}
&= \int_{B(0,\epsilon)^+} \eta_\epsilon(y)\langle x, y\rangle \; dy + \int_{B(0,\epsilon)^+} \eta_\epsilon(-y)\langle x, -y\rangle \; dy \\
&= \int_{B(0,\epsilon)^+} \eta_\epsilon(y)\langle x, y\rangle \; dy + \int_{B(0,\epsilon)^+} \eta_\epsilon(y)\langle x, -y\rangle \; dy \\
&= 0, \qquad (1.14)
\end{aligned}$$

where we used the fact in the last line that η_ϵ is radial.

The last integral in 1.13 is a finite constant:

$$\int_{B(0,\epsilon)} \eta_\epsilon(y) \mid y \mid^2 \ dy \leq \int_{B(0,\epsilon)} \frac{1}{e}\epsilon^2 \ dy = \frac{\alpha(n)\epsilon^{n+2}}{e} < \infty.$$

Because the last term in 1.13 is constant, it does not affect the derivatives of the whole sum, and it thus has no effect on convexity. Now since $\phi(x) + \frac{\lambda}{2} \mid x \mid^2$ is convex, $\phi_\epsilon(x) + \frac{\lambda}{2} \mid x \mid^2$ is convex. That is, $\phi_\epsilon(x)$ is semiconvex, with convexity constant λ.

Therefore each of the sets

$$K_\epsilon = \left\{x \in \overline{B}(\hat{x}, r); \phi_\epsilon(x) + \langle p, x\rangle \text{ has a max at } x \text{ for some } p \in B_\delta\right\}$$

satisfies the property proved above for C^2 functions:

$$\mathcal{L}^n(K_\epsilon) \geq \frac{\mathcal{L}^n(B_\delta)}{\mid \lambda \mid^n} > 0.$$

Let $U_m = \bigcup_{l=m}^\infty K_{1/l}$, $g_m = \chi_{U_m}$, and $U = \bigcap_{m=1}^\infty \bigcup_{l=m}^\infty K_{1/l}$. Since $U_1 \subset U_2 \subset ...$, we see that g_m decreases monotonically to χ_U. Also note that $g_1 \in L^1(\mathbf{\Omega})$. Using Lebesgue's Monotone Convergence Theorem, we get

$$\mathcal{L}^n(U) = \int_{\mathbf{\Omega}} \chi_U \ dx = \lim_{m\to\infty} \int_{\mathbf{\Omega}} g_m \ dx \geq \frac{\mathcal{L}^n(B_\delta)}{\mid \lambda \mid^n} > 0.$$

That is, $\mathcal{L}^n(\bigcap_{m=1}^\infty \bigcup_{l=m}^\infty K_{1/l}) \geq \frac{\mathcal{L}^n(B_\delta)}{|\lambda|^n} > 0$.

Last, we wish to show that $\bigcap_{m=1}^\infty \bigcup_{l=m}^\infty K_{1/l} \subset K$. Suppose $x \in \bigcap_{m=1}^\infty \bigcup_{l=m}^\infty K_{1/l}$. Then for infinitely many $m \in \mathbb{N}$, there exists $p = p(m) \in B_\delta$ such that $(\phi_p(m))_{1/m}$ has a max at x. Observe that $\{p(m)\}_{m=1}^\infty$ has an accumulation point, call it p', in B_δ. Hence there is a subsequence $\{m_i\}_{i=1}^\infty \subset \mathbb{N}$ such that $(\phi_{p(m_i)})_{1/m_i}$ has a max at x for each $i \in \mathbb{N}$ and $p(m_i) \to p'$ as $i \to \infty$.

Note that

$$\int_{B(0,\frac{1}{m_i})} \eta_{\frac{1}{m_i}} \langle p(m_i), x-y\rangle \; dy = \langle p(m_i), x\rangle - \int_{B(0,\frac{1}{m_i})} \eta_{1/m_i}(y)\langle p(m_i), y\rangle \; dy$$

and

$$\int_{B(0,1/m_i)} \eta_{1/m_i}(y)\langle p(m_i), y\rangle \; dy = 0,$$

by the same kind of argument as in 1.14.

Therefore $(\phi_{p(m_i)})_{1/m_i}$ converges uniformly to $\phi_{p'}$, and $\phi_{p'}$ then has a minimum at x. So x satisfies the required properties for inclusion in K.

That is, $\bigcap_{m=1}^{\infty}\bigcup_{l=m}^{\infty} K_{1/l} \subset K$. Hence

$$0 < \mathcal{L}^n(\bigcap_{m=1}^{\infty}\bigcup_{l=m}^{\infty} K_{1/l}) \leq \mathcal{L}^n(K).$$

■

Chapter 2

Viscosity Solutions

In this chapter, we lay out the concept of viscosity solutions for second order PDE. We would like to be able to solve a wide class of PDE of the form

$$F(x, u(x), Du(x), D^2u(x)) = 0,$$

where F is some function that maps $\mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times S(n) \to \mathbb{R}$. Here $S(n)$ is the set of $n \times n$ symmetric matrices with real entries.

The problem is that there aren't enough twice-differentiable functions to provide solutions for most nonlinear F, nor even linear ones with nonsmooth coefficients. So to develop an interesting theory and enough solutions to satisfy applications, we need to generalize what we mean by a "solution" of a PDE in such a way that allows for nonsmooth functions to be considered solutions, provided they meet certain appropriate requirements. Defining what those requirements should be is the point of this chapter.

In the theory of linear PDE, one can talk about Sobolev functions or distributions, because they make sense under linear operations. Thus, one can define a solution of a linear PDE to be a weakly-differentiable function that satisfies the given PDE in the sense of distributions.

But in the nonlinear situation, this method breaks down. For example, if the PDE involves taking the square of a function or its derivatives, that becomes a difficulty. Distributions are essentially objects that live only in a linear space. Although we can add distributions and multiply them by C^∞ functions, it does not in general make sense to talk about multiplying two distributions.

So we need to develop other frameworks for dealing with nonlinear PDE. For example, in the situation of Hamilton-Jacobi equations, the Hopf-Lax formula demonstrates such a method: it produces solutions by minimizing some functional related to the PDE. But methods like this still require that an appropriate and useful meaning be assigned to the word "solution."

A step in this direction is defining so-called *viscosity solutions* of second order PDE. This is a generalization of the classical theory of subharmonic functions of F. Riesz and T. Rado. The modern aspects of this theory are due to Lions, Crandall, Ishii, Evans and Jensen.

The adjective "viscosity" is a relic of an earlier theory out of which this one grew. It was a theory for first-order PDE that worked as follows: One could approximate a fully-nonlinear first-order PDE with a quasilinear second order PDE whose solutions were easier to define (they were always smooth). The approximation was achieved in part by adding a small "viscosity" term – a constant multiple of the Laplacian of a solution – to the PDE. As this constant multiple was sent to zero, the original PDE would emerge. This was called the "method of vanishing viscosity." (For more on the first-order theory, see [Evans, 1998]) This first-order viscosity-solution theory also makes rigorous the Hopf-Lax formula described above as a "solution" to a PDE.

The definition of "viscosity solution" that was developed with that theory evolved into the one we turn to now for second-order PDE.

2.1 "Proper" Functions

Second order PDE are those that can be written as

$$F(x, u(x), Du(x), D^2u(x)) = 0,$$

for some function F. There is no general theory for the solution of all PDE, so we must restrict our attention to particular classes of functions F.

In this essay, we will be concerned with *proper* functions F.

Definition 2.1 *A function* $F : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \times S(n) \to \mathbb{R}$ *is said to be* proper *if it satisfies*

$$F(x,r,p,X) \leq F(x,s,p,X) \ \ \text{whenever } r \leq s, \qquad \text{and}$$

$$F(x,r,p,X) \leq F(x,r,p,Y) \ \ \text{whenever } Y \leq X.$$

So "proper" just means that F is increasing in the second variable, r, and decreasing in the last variable, X. The ordering on matrices in $S(n)$ is given by

$$Y \leq X \Leftrightarrow Y - X \leq 0,$$

that is to say, $Y \leq X$ whenever the matrix $Y - X$ is negative semidefinite.

EXAMPLE 1: Let the symmetric matrix $\{A_{i,j}(x)\}$ be positive semidefinite for each x and let $c(x) \geq 0$. Then we can write the elliptic PDE in non-divergence form

$$-\sum_{i,j=1}^{n} A_{i,j}(x)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i} + c(x)u(x) = f(x),$$

or we can write the PDE as $F(x, u(x), Du(x), D^2u(x)) = 0$ by setting

$$F(x,r,p,X) = -\text{trace}(A(x)X) + \sum_{i=1}^{n} b_i(x)p_i + c(x)r - f(x).$$

With these conditions on $A(x)$ and $c(x)$, F is proper.

EXAMPLE 2: Suppose $F(x, u(x), Du(x))$ is a first order equation, so that it does not depend on D^2u. Then F is proper as long as it is increasing in the second variable. For instance,

$$F(x,r,p) = f(x,p) + g(x)r$$

is a proper function as long as $g(x) \geq 0$ for all x. This corresponds to the first order PDE

$$f(x, Du(x)) + g(x)u(x) = 0.$$

EXAMPLE 3: Consider the p-Laplace equation $\Delta_p u(x) = f(x)$ defined by $\Delta_p u(x) = \operatorname{div}(|Du|^{p-2})Du)$. This is a nonlinear second order equation in general, though in the case $p = 2$ it simplifies to the regular (linear) Laplace equation. Differentiation shows that

$$\Delta_p u = \sum_{i=1}^{n} \left((p-2)|Du|^{p-4} + |Du|^{p-2}\right) \frac{\partial^2 u}{\partial x_i^2}.$$

Therfore the p-Laplace equation can be written as above using the proper function

$$F(x, r, q, Y) = \sum_{i=1}^{n} \left((p-2)|q|^{p-4} + |q|^{p-2}\right) Y_{i,i} - f(x).$$

These are simple examples that are by no means exhaustive. But they do illustrate the fact that we have plenty of PDE that can be represented by proper functions F. For more examples, see [CIL].

Through the rest of this essay, F will be a proper function as defined above. We also assume that F is continuous from now on.

2.2 Subjets and Superjets

In order to think of a nonsmooth function as a solution to a PDE, we must generalize the idea of derivative. This can be done globally, as in the case of Sobolev functions, but to work with nonlinear equations, we need a local point of view – we need to be able to discuss pointwise values so we can perform nonlinear operations on these generalized derivatives.

For the moment, suppose $u \in C^2(\mathbf{\Omega})$. Then we call u a classical subsolution of the PDE given by $F = 0$ if $F(x, u(x)Du(x), D^2u(x)) \leq 0$ for all $x \in \mathbf{\Omega}$. Suppose that $\phi \in C^2(\mathbf{\Omega})$ and $\tilde{x}$ is a local maximum of $u - \phi$. Then $D(u - \phi)(\tilde{x}) = 0$, so $Du(\tilde{x}) = D\phi(\tilde{x})$. Similarly, $D^2(u - \phi)(\tilde{x}) \leq 0$, so $D^2u(\tilde{x}) \leq D^2\phi(\tilde{x})$. Now because we assume F to be proper, we get

$$\begin{aligned} F(\tilde{x}, u(\tilde{x}), D\phi(\tilde{x}), D^2\phi(\tilde{x})) &= F(\tilde{x}, u(\tilde{x}), Du(\tilde{x}), D^2\phi(\tilde{x})) \\ &\leq F(\tilde{x}, u(\tilde{x}), Du(\tilde{x}), D^2u(\tilde{x})) \\ &\leq 0. \end{aligned}$$

Notice that the inequality $F(\tilde{x}, u(\tilde{x}), D\phi(\tilde{x}), D^2\phi(\tilde{x})) \leq 0$ does not depend on the derivatives of u, but that it does hold for any C^2 function ϕ and local maximum $\tilde{x}$ of $u - \phi$. This observation leads directly to our notion of generalized derivative.

We look for vectors $p \in R^n$ and symmetric matrices $X \in S(n)$ that can be used in place of the first and second derivatives in the function F. If we also see that, near this local maximum $\tilde{x}$ of $u - \phi$, we have $u(x) \leq u(\tilde{x}) - \phi(\tilde{x}) + \phi(x)$, then we use a Taylor approximation to write

$$u(x) \leq u(\tilde{x}) + \langle D\phi, x - \tilde{x}\rangle + \frac{1}{2}\langle D^2\phi(x - \tilde{x}), x - \tilde{x}\rangle + o(|\ x - \tilde{x}\ |^2) \ \text{ as } x \to \tilde{x}. \quad (2.1)$$

This notation might require explanation. We write $f(z) \leq g(z) + o(z^2) \quad as\ z \to w$ to mean that there exists a function $h(z) \geq 0$ such that $h(z) = o(z^2) \quad as\ z \to w$ and $f(z) = g(z) + h(z)$ in a neighborhood of w. A similar definition exists for $f(z) \geq g(z) + o(z^2)$, this time with $h(z) \leq 0$.

In general, equation 2.1 may be true when we replace $D\phi(\tilde{x})$ with some other vectors and $D^2\phi(\tilde{x})$ with some other matrices. For instance, suppose $f(z) = z^2$ is defined on the interval $(-1, 1)$. Then we can write

$$f(z) = f(0) + \langle 0, z\rangle + \frac{1}{2}\langle 2z, z\rangle.$$

Using less than the full strength of the equality here, we can write it as an inequality,

$$f(z) \leq f(0) + \langle p, z\rangle + \frac{1}{2}\langle Xz, z\rangle,$$

where we can take $p = 0$ and $X = 2$. But notice that we could also take, say, $X = 4$, and this inequality will still be true.

Or instead, if we look at $f(z) = \mid z \mid$ on the interval $(-1, 1)$, then we could take $p = -1$. But we couldn't find an X to work.

We can think of the situation like this: suppose we throw out all the terms of f with order higher than 2. Then we can take p and X to be the first and second derivatives, respectively, of any quadratic that touches the graph of f from above at $z = w$. These are the vectors and matrices that satisfy 2.1. (And we can think of these quadratics as the C^2 functions ϕ, locally and modulo their higher order terms, that give a local maximum of $(u - \phi)$ at the desired point.)

We return to the notation of Taylor approximations and denote the set of pairs (p, X) that work in this way as follows.

Definition 2.2 *For a function* $u : \mathbf{\Omega} \to \mathbb{R}$, *we define the* second order superjet of u at x *by*

$$J^{2,+}u(x) = \{(p, X) \in \mathbb{R}^n \times S(n) \mid u(y) \leq u(x) + \langle p, y - x\rangle + \frac{1}{2}\langle X(y - x), y - x\rangle + o(\mid y - x \mid^2)\}. \tag{2.2}$$

Similarly, we define the second order subjet of u at x *by*

$$J^{2,-}u(x) = \{(p, X) \in \mathbb{R}^n \times S(n) \mid u(y) \geq u(x) + \langle p, y - x\rangle + \frac{1}{2}\langle X(y - x), y - x)\rangle + o(\mid y - x \mid^2)\}. \tag{2.3}$$

Last we define $J^2 u(x) = J^{2,+} \bigcap J^{2,-}u(x)$.

Note for later reference that $J^{2,+}u(x) = -J^{2,-}(-u)(x)$, and that the intersection $J^2 u(x)$ of the subjet and superjet of u at x is nonempty if and only if u is twice-differentiable at x (that is, there exists a second-order Taylor approximation to u).

In summary, looking at the Taylor-expansion properties of differentiable functions leads to our generalized pointwise derivatives – the jets – just as looking at the integration-by-parts properties of differentiable functions leads to distributional derivatives. Taylor expansions are local events, whereas integration is a more global event, and these differences lead to greater variety in the applications we can now approach.

2.3 Viscosity Solutions

Definition 2.3 *A function $u \in USC(\mathbf{\Omega})$ is called a* viscosity subsolution *of*

$$F(x, u(x), Du(x), D^2u(x)) = 0 \tag{2.4}$$

if

$$F(x, u(x), p, X) \leq 0 \text{ for all } x \in \mathbf{\Omega} \text{ and all } (p, X) \in J^{2,+}u(x).$$

A function $v \in LSC(\mathbf{\Omega})$ is called a viscosity supersolution *of equation 2.4 if*

$$F(x, v(x), p, X) \geq 0 \text{ for all } x \in \mathbf{\Omega} \text{ and all } (p, X) \in J^{2,-}v(x).$$

A function is a viscosity solution *of equation 2.4 if it is both a viscosity subsolution and a viscosity supersolution.*

This definition is perfectly consistent with the classical definition of a solution, because the intersection of the subjet and the superjet of a twice differentiable function contains just the first and second derivatives of the function, and these inequalities are satisfied immediately.

From here on we will usually drop the term "viscosity" and simply refer to solutions, subsolutions and supersolutions for the rest of this essay.

Definition 2.3 imposes conditions of semicontinuity on the subsolutions and supersolutions. This forces a solution to be continuous, since upper-semicontinuity plus lower-semicontinuity implies continuity.

One reason for requiring semicontinuity is that upper-semicontinuous functions achieve their local maxima, so it is possible to find local maxima of $u - \phi$, as discussed in Section 2, for subsolutions u and C^2 functions ϕ. It is similarly possible to find local maxima for $\phi - v$, when v is a lower-semicontinuous supersolution.

2.4 Closures of Jets

In our later developments, we will approximate an interior maximum of $u-v$, with u a subsolution and v a supersolution, by considering a sequence of points x_m, y_m in $\mathbf{\Omega}$ chosen such that $u(x_m) - v(y_m)$ approaches this internal maximum. We will show that points in these sequences have nonempty "jet closures." Even though the jets themselves might be empty at the desired max of $u-v$ and along the approximating sequences, we will be able to use the inequalities satisfied by subsolutions and supersolutions thanks to the continuity of the function F. This idea will be made more clear after the next definition.

To facilitate our work in this direction, we use the following notation.

Definition 2.4 *The* closure *of the subjet $J^{2,+}_{\mathbf{\Omega}}u(x)$ is defined to be*

$$\overline{J}^{2,+}_{\mathbf{\Omega}}u(x) = \{(p,X) \mid (x_m, u(x_m), p_m, X_m) \to (x, u(x), p, X) \quad \text{as } m \to \infty \text{ and } (p_m, X_m) \in J^{2,+}_{\mathbf{\Omega}}u(x_m)\}. \tag{2.5}$$

Similarly, the closure *of the superjet $J^{2,-}_{\mathbf{\Omega}}u(x)$ is*

$$\overline{J}^{2,-}_{\mathbf{\Omega}}u(x) = \{(p,X) \mid (x_m, u(x_m), p_m, X_m) \to (x, u(x), p, X) \quad \text{as } m \to \infty \text{ and } (p_m, X_m) \in J^{2,-}_{\mathbf{\Omega}}u(x_m)\}. \tag{2.6}$$

Finally, as might be expected, we set

$$\overline{J}^{2}_{\mathbf{\Omega}}u(x) = \{(p,X) \mid (x_m, u(x_m), p_m, X_m) \to (x, u(x), p, X) \quad \text{as } m \to \infty \text{ and } (p_m, X_m) \in J^{2}_{\mathbf{\Omega}}u(x_m)\}. \tag{2.7}$$

These definitions will provide a useful notational convenience later on.

Suppose u is a viscosity subsolution of $F = 0$ and $(p, X) \in \overline{J}^{2,+}_{\mathbf{\Omega}}u(x)$ for some x in the domain $\mathbf{\Omega}$. Then by the definition of the closure of a superjet, there exist $(p_m, X_m) \in J^{2,+}_{\mathbf{\Omega}}u(x_m)$ with $(x_m, u(x_m), p_m, X_m) \to (x, u(x), p, X)$. Since u is a subsolution, we have

$$F(x_m, u(x_m), p_m, X_m) \leq 0.$$

Because F is continuous, we can take the limit as $m \to \infty$ to obtain

$$F(x, u(x), p, X) \leq 0.$$

So even though (p, X) is not in an appropriate jet as required by the definition of a viscosity subsolution, and the genuine jet may in fact be empty, we can still write this as a meaningful inequality for a pair (p, X) in the closure of the superjet.

Similar statements hold for supersolutions and the closures of subjets.

Chapter 3

The Comparison Principle

In this chapter, we first work to prove that the closures of jets are nonempty at local maxima of upper semicontinuous functions. At the end of the chapter, we will use this fact to prove the uniqueness of solutions to the Dirichlet problem for certain PDE, when solutions are understood in the viscosity sense. In between, we will cover the "comparison principle" which states that subsolutions are less than or equal to supersolutions throughout a domain on which the supersolution's boundary values majorize those of the subsolution. The uniqueness theorem is then an immediate consequence of this comparison principle.

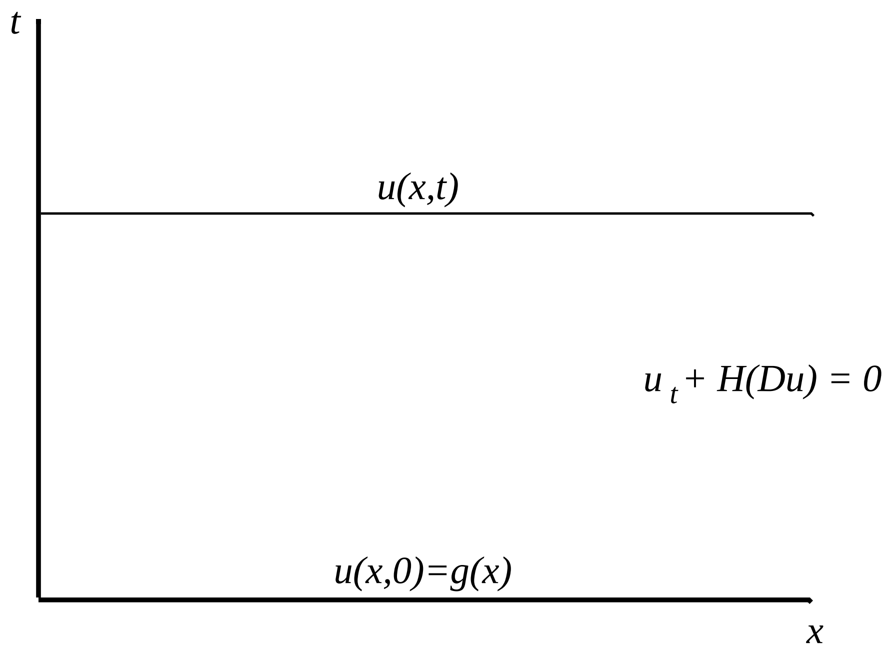
But before we start any of this, let us first try to motivate the primary tool used in this chapter, the so-called "sup convolutions" which are used to obtain the nonempty jets we need in the maximum principle.

3.1 Hamilton-Jacobi Equations

Let us briefly recall the Hamilton-Jacobi equation:

$$\begin{cases} u_t + H(Du) = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases} \tag{3.1}$$

Here, $H : \mathbb{R}^n \to \mathbb{R}$ is a given convex function (called the Hamiltonian) that is unbounded in all directions, and $u_t + H(Du) = 0$ is the differential equation we wish to solve, subject to the initial conditions that a solution u must equal the given function g on the initial boundary, $\mathbb{R}^n \times \{t = 0\}$.



This Hamilton-Jacobi equation is a first-order nonlinear equation in general, and it can be thought of as taking an initial function and evolving it in time according to the differential equation $u_t + H(Du) = 0$.

One can find intimate connections between viscosity solutions and Hamilton-Jacobi equations. The most basic connection is that the first-order theory of viscosity solutions was developed to make rigorous the notion of a solution to 3.1. In general, the solutions of 3.1 are not smooth for all times t. So we need to be careful and define what we mean by a generalized solution to the PDE. Distributions are no good in this setting, because they do not ordinarily make sense under nonlinear operations (for example, you cannot multiply distributions together, but we may very well have a given Hamiltonian function H that performs some such operation on the components of Du).

The first-order theory of viscosity solutions provided a framework for solutions to 3.1 by making rigorous the "method of vanishing viscosity" that was sometimes used to solve these PDE. (See [E] for more details on this, as well as all the other facts mentioned in this section.)

An explicit formula for solving Hamilton-Jacobi equations which is also made rigorous by the theory of first-order viscosity solutions is the so-called Hopf-lax formula. To a given Hamiltonian function H we associate a Lagrangian function L (also called the "convex dual" of H) defined by

$$L(p) = \sup_{q \in \mathbb{R}^n} \{p \cdot q - H(q)\} \qquad \text{for } p \in \mathbb{R}^n. \tag{3.2}$$

Using the calculus of variations, one can then prove:

Theorem 3.1 ***(Hopf-Lax formula)*** *If $x \in \mathbb{R}^n$ and $t > 0$, then the solution of the Hamilton-Jacobi equation 3.1 in the (first-order) viscosity sense is*

$$u(x,t) = \inf_{y \in \mathbb{R}^n} \left\{ tL\left(\frac{x-y}{t}\right) + g(y) \right\}. \tag{3.3}$$

EXAMPLE: Suppose we want to solve the PDE

$$\begin{cases} u_t + \frac{1}{2}|Du|^2 = 0 & \text{in } \mathbb{R}^n \times \mathbb{R} \\ u = g & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

This is clearly a Hamilton-Jacobi equation with Hamiltonian $H(p) = \frac{1}{2}|p|^2$.

Since $q \cdot p - \frac{1}{2}|p|^2$ is differentiable, we can use calculus to find a solution of 3.2. Doing so gives us the Lagrangian

$$L(q) = \frac{1}{2}|q|^2 \ (= H(q)).$$

Therefore, using the Hopf-Lax Formula 3.3, we obtain the solution

$$u(x,t) = \inf_{y \in \mathbb{R}^n} \left\{ \frac{1}{2t}|x-y|^2 + g(y) \right\}. \tag{3.4}$$

Next, we mention some of the properties of solutions defined by the Hopf-Lax formula. But first, another definition is in order.

Definition 3.2 *A function $f : \mathbf{\Omega} \to \mathbb{R}$ is said to be* concave *on the domain $\mathbf{\Omega}$ if $-f$ is convex on $\mathbf{\Omega}$. A function f is said to be* semiconcave *on $\mathbf{\Omega}$ if $-f$ is semiconvex on $\mathbf{\Omega}$.*

It is apparent that a function f is semiconcave if and only if there exists $\lambda > 0$ such that $f(x) - \frac{\lambda}{2} \mid x \mid^2$ is concave. This λ is called a *semiconcavity constant*, and we can see that it is also a semiconvexity constant for $-f$. We bring up this definition because the properties of the Hopf-Lax formula are most easily described in this way.

Suppose that the following are true: (1) $H : \mathbb{R}^n \to \mathbb{R}$ is a smooth Hamiltonian, and (2) $g : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz-continuous initial-value function. Then the function u defined by the Hopf-Lax formula is Lipschitz continuous in $\mathbb{R}^n \times [0, \infty)$. By Rademacher's Theorem, we immediately know that u is differentiable almost everywhere.

But there's more. The solution u is semiconcave if g is semiconcave or if the Hamiltonian H is "uniformly convex," which means simply that there is a constant $\theta > 0$ such that

$$\sum_{i,j=1}^{n} H_{p_i p_j}(p)\xi_i \xi_j \geq \theta |\xi|^2 \qquad \text{for all } p, \xi \in \mathbb{R}^n.$$

For instance, the Hamiltonian in our last example, $H(p) = \frac{1}{2}|p|^2$, is uniformly convex with constant $\theta = \frac{1}{2}$.

So nice Hamiltonians, such as the one in our example, have a regularizing effect on solutions. Given a Lipschitz initial-value function, we can evolve it in time with a Hamilton-Jacobi equation, and at every positive time t, the solution will be semiconcave, which makes it the negative of a semiconvex function. (We will prove these convexity properties only in the special case of the previous example; for a more general approach, the reader may consult [E].) Now recall Aleksandrov's Theorem from Chapter 1: (semi)convex functions are *twice-differentiable* almost everywhere. So our solution must be twice-differentiable almost everywhere, too.

All this is to say that a Hamilton-Jacobi equation, with a uniformly convex Hamiltonian, will evolve a Lipschitz (read: once-differentiable a.e.) function into a semiconcave (read: twice-differentiable a.e.) function.

Realizing what is going on here, and looking to associate nonempty jets to given functions as we will try to do later, we might wonder the following: *is it possible to evolve an upper-semicontinuous function in a similar manner with a Hamilton-Jacobi equation to get a solution that is twice-differentiable almost everywhere? And if so, will it tell us anything about the regularity properties of the initial function?*

The answers to both of these questions are "Yes!"

3.2 Sup Convolutions

Look again at the Hopf-Lax formula 3.3. This expression can also be written as a supremum by using the identity $-(\inf A) = \sup(-A)$, for any set A of real numbers, yielding

$$-u(x,t) = \sup_{y\in\mathbb{R}^n}\left\{-g(y) - \frac{1}{2t}|x-y|^2\right\}.$$

Let us clean up the notation a bit by writing $w(x,\lambda) = -u(x,\frac{1}{\lambda})$ and $h(y) = -g(y)$. Now we are looking at

$$w(x,\lambda) = \sup_{y\in\mathbb{R}^n}\left\{h(y) - \frac{\lambda}{2}|x-y|^2\right\}. \tag{3.5}$$

Despite the fact that we are not assuming g to be Lipschitz, we can still show that, under lesser conditions on g, u is a semiconcave function of x for all fixed $t > 0$. (We will prove this shortly.)

In other words, given an upper-semicontinuous function h, the function $w(x,\lambda)$ defined by 3.5 is semiconvex in x. This result is what we need to proceed: we will find jets on the function $w(x,\lambda)$ using Aleksandrov's Theorem, and somehow we will show that they imply the existence of jets on the function h.

It should now be clear that we will make thorough use of the following definition.

Definition 3.3 *Suppose $f \in USC(\mathbb{R}^m)$ is bounded above. For $\lambda > 0$, define the* sup convolution f_λ *of f by*

$$f_\lambda(x) = \sup_{\xi\in\mathbb{R}^m}\left(f(\xi) - \frac{\lambda}{2} \mid x-\xi \mid^2\right). \tag{3.6}$$

The next lemma tells us of the the existence of second-order jets with good bounds for semiconvex functions. As the reader might anticipate, we intend to apply this result to the sup convolutions as an intermediate step to deriving the existence of jets on the original function.

Lemma 3.4 *If $f \in C(\mathbb{R}^m)$, $B \in S(m)$, $f(\xi) + \frac{\lambda}{2} \mid \xi \mid^2$ is convex and*

$$\max_{\mathbb{R}^m}(f(\xi) - \frac{1}{2}\langle B\xi,\xi\rangle) = f(0),$$

then there is an $X \in S(m)$ *such that* $(0, X) \in \overline{J}^2 f(0)$ *and* $-\lambda \leq X \leq B$.

Proof: Since $\max_{\mathbb{R}^m}(f(\xi) - \frac{1}{2}\langle B\xi, \xi\rangle) = f(0)$, the function

$$g(\xi) = f(\xi) - \frac{1}{2}\langle B\xi, \xi\rangle - \mid \xi \mid^4$$

has a strict local max at $\xi = 0$. By Aleksandrov's Theorem, $f(\xi) + \frac{\lambda}{2} \mid \xi \mid^2$ is twice differentiable almost everywhere (because it is convex). Hence $f(\xi)$ is twice differentiable almost everywhere, and so is $g(\xi)$.

By Jensen's Lemma, for every $\delta > 0$, there is a set $S \subset \overline{B}_\delta$ such that S has positive measure, and for any $\xi_\delta \in S$ there exists $q_\delta \in \overline{B}_\delta$ for which

$$g(\xi) + \langle q_\delta, \xi\rangle \tag{3.7}$$

has a max at ξ_δ. Since $g(\xi)$ is twice differentiable almost everywhere, it follows that we can choose ξ_δ so $Dg(\xi_\delta)$ and $D^2 g(\xi_\delta)$ exist.

Observe that $Dg(\xi_\delta) + q_\delta = 0$, because ξ_δ is a local max of $g(\xi) + \langle q_\delta, \xi\rangle$. So we have

$$\begin{aligned}
\mid Df(\xi_\delta) \mid &\leq \mid Dg(\xi_\delta) \mid + \mid B\xi_\delta \mid + 4 \mid \xi_\delta \mid^3 \\
&\leq \mid -q_\delta \mid + \parallel B \parallel \mid \xi_\delta \mid + 4 \mid \xi_\delta \mid^3 \\
&\leq \mid \delta \mid + \parallel B \parallel \mid \delta \mid + 4 \mid \delta \mid^3 \\
&= O(\delta) \quad \text{as } \delta \to 0^+.
\end{aligned}$$

Since $f(\xi) + \frac{\lambda}{2} \mid \xi \mid^2$ is convex, it has a nonnegative second derivative almost everywhere:

$$D^2 f(\xi) + \lambda I \geq 0,$$

that is,

$$-\lambda I \leq D^2 f(\xi) \qquad \text{a.e.} \tag{3.8}$$

Moreoever, at ξ_δ, $g(\cdot) + \langle q_\delta, \cdot \rangle$ has a local max, so $D^2 g$ is nonpositive:

$$D^2 g(\xi_\delta) = D^2 f(\xi_\delta) - B - 4 \mid \xi_\delta \mid^2 I + 8\xi_\delta \xi_\delta^T \leq 0.$$

Thus, it follows that

$$D^2 f(\xi_\delta) \leq B + O(\delta^2) I. \tag{3.9}$$

Combining 3.8 with 3.9 yields

$$-\lambda I \leq D^2 f(\xi_\delta) \leq B + O(\delta^2) I. \tag{3.10}$$

Since ξ_δ is chosen so that f is twice differentiable there, we have

$$(Df(\xi_\delta), D^2 f(\xi_\delta)) \in J^2 f(\xi_\delta).$$

Choose a sequence $\{\delta_k\}_{k=1}^\infty$ so that $\delta_k \to 0$ as $k \to \infty$, and $\mid \delta_k \mid < 1$ for all $k \in \mathbb{N}$. Then

$$(Df(\xi_{\delta_k}), D^2 f(\xi_{\delta_k})) \in J^2 f(\xi_{\delta_k}),$$

and provided the limits exist as $k \to \infty$, we see

$$(\lim_{k\to\infty} Df(\xi_{\delta_k}), \lim_{k\to\infty} D^2 f(\xi_{\delta_k})) \in \overline{J}^2 f(0).$$

Since $\mid Df(\xi_{\delta_k}) \mid = O(\delta_k)$ as $k \to \infty$, we have

$$\lim_{k\to\infty} Df(\xi_{\delta_k}) = 0.$$

Put $E = \{M \in S(m) \mid -\lambda I \leq M \leq B + I\}$. Then for $M \in E$,

$$\mid M \mid \leq \max\{\mid -\lambda I \mid, \mid B + I \mid\}.$$

This shows that E is a bounded subset of the Euclidean space $\mathbb{R}^{m^2}$. Clearly E is closed in $\mathbb{R}^{m^2}$. Thus E is compact. Infinite sequences in compact sets have convergent subsequences, so we can pass to a subsequence of $\{\delta_k\}_{k=1}^{\infty}$ such that $D^2 f(\xi_{\delta_k})$ converges to some $X \in S(m)$. This is the matrix X we needed: $(0, X) \in \overline{J}^2 f(0)$.

Finally, by 3.10, since $-\lambda I \leq X \leq B + O(\delta_k)I$ for all δ_k in the chosen subsequence, and $\delta_k \to 0$, we get $-\lambda I \leq X \leq B$.

■

Lemma 3.5 *Let $\lambda > 0$, $f \in USC(\mathbb{R}^n)$ be bounded above globally, and*

$$f_\lambda(x) = \sup_{y \in \mathbb{R}^n} \{f(y) - \frac{\lambda}{2}|x - y|^2\}$$

be the sup convolution of f. If $(q_0, Y_0) \in J^{2,+} f_\lambda(x_0)$, then $(q_0, Y_0) \in J^{2,+} f(x_0 + \frac{q_0}{\lambda})$ and $f_\lambda(x_0) + \frac{1}{2\lambda}|q_0|^2 = f(x_0 + \frac{q_0}{\lambda})$.

Proof: Suppose $(q_0, Y_0) \in J^{2,+} f_\lambda(x_0)$. Since f is upper-semicontinuous we can find $y_0 \in \mathbb{R}^n$ such that the supremum in the sup convolution is obtained, i.e.

$$f_\lambda(x_0) = f(y_0) - \frac{\lambda}{2}|x_0 - y_0|^2. \tag{3.11}$$

<u>Claim</u>: For any $x, y \in \mathbb{R}^n$, we have

$$\begin{aligned} f(x) - \frac{\lambda}{2}|y - x|^2 \leq f(y_0) + \frac{\lambda}{2}|y_0 - x_0|^2 + \langle q_0, y - x_0 \rangle \\ + \frac{1}{2}\langle Y_0(y - x_0), y - x_0 \rangle + o(|y - x_0|^2). \end{aligned} \tag{3.12}$$

In particular, this is true for $y = x - y_0 + x_0$, and making this substitution in 3.12 yields

$$\begin{aligned} f(x) + \frac{\lambda}{2}|x_0 - y_0|^2 \leq f(y_0) + \frac{\lambda}{2}|y_0 - x_0|^2 + \langle q_0, x - y_0 \rangle \\ + \frac{1}{2}\langle Y_0(x - y_0), x - y_0 \rangle + o(|x - y_0|^2). \end{aligned}$$

Now cancel the term $\frac{\lambda}{2}|x_0 - y_0|^2$ from each side to obtain

$$f(x) \leq f(y_0) + \langle q_0, x - y_0 \rangle + \frac{1}{2}\langle Y_0(x - y_0), x - y_0 \rangle = o(|x - y_0|^2).$$

This holds for all $x \in \mathbb{R}^n$, and it therefore says that

$$(q_0, Y_0) \in J^{2,+} f(y_0). \tag{3.13}$$

That is to say, because the sup convolution has a non-empty superjet at a point x_0, the original function must also have a non-empty superjet (and it contains the same pair!) at a point y_0 that depends on x_0. Next we want to demonstrate that this y_0 is close to x_0.

We can weaken 3.12 by consuming the last inner product into the error term and writing

$$f(x) - \frac{\lambda}{2}|y - x|^2 \leq f(y_0) + \frac{\lambda}{2}|y_0 - x_0|^2 + \langle q_0, y - x_0 \rangle + O(|y - x_0|^2). \tag{3.14}$$

Again, this holds for all $x, y \in \mathbb{R}^n$, and now we make the substitutions $x = y_0$ and $y = x_0 + \alpha(\lambda(x_0 - y_0) + q_0)$, where α is a negative number of small magnitude (whose relevance will be made apparent in a moment). Make this substitution in 3.14 to see that

$$\begin{aligned} f(y_0) - &\frac{\lambda}{2}|x_0 - y_0 + \alpha(\lambda(x_0 - y_0) + q_0)|^2 \\ &\leq f(y_0) + \frac{\lambda}{2}|y_0 - x_0|^2 + \langle q_0, \alpha(\lambda(x_0 - y_0) + q_0) \rangle + O(|\alpha(\lambda(x_0 - y_0) = q_0)|^2). \end{aligned}$$

Cancel $f(y_0)$ from both sides and expand the squared norm on the left to get

$$\begin{aligned} -\frac{\lambda}{2}|x_0 - y_0|^2 - \lambda\langle x_0 - y_0, \alpha(\lambda(x_0 - y_0) + q_0) \rangle - \frac{\lambda}{2}|\alpha(\lambda(x_0 - y_0) + q - 0)|^2 \\ \leq -\frac{\lambda}{2}|y_0 - x_0|^2 + \langle q_0, \alpha(\lambda(x_0 - y_0) + q_0) \rangle + O(|\alpha(\lambda(x_0 - y_0) + q_0)|^2). \end{aligned}$$

Cancel like terms from each side again, and consume higher-order terms into the error:

$$-\lambda\langle x_0 - y_0, \alpha(\lambda(x_0 - y_0) + q_0)\rangle \leq \langle q_0, \alpha(\lambda(x_0 - y_0) + q_0)\rangle + O(|\alpha(\lambda(x_0 - y_0) + q_0)|^2).$$

Finally, we move everything to the right side of the equation and combine the two inner products into one, then factor the α out of the inner product to reach

$$\begin{aligned} 0 &\leq \langle q_0 + \lambda(x_0 - y_0), \alpha, \alpha(\lambda(x_0 - y_0) + q_0)\rangle + O(|\alpha(\lambda(x_0 - y_0) + q_0)|^2) \\ &\leq \alpha|\lambda(x_0 - y_0) + q_0)|^2 + O(|\alpha(\lambda(x_0 - y_0) + q_0)|^2). \end{aligned} \tag{3.15}$$

Now remember that x_0, y_0 and q_0 were fixed vectors in $\mathbb{R}^n$, so the expression in 3.15 really only depends on α. The first term on the right is linear in α, and the error term can be written as $O(|\alpha|^2)$. This is where our choice of $\alpha < 0$ becomes useful: we can drop the error term if the magnitude of α is sufficiently small, leaving the inequality

$$0 \leq \alpha|\lambda(x_0 - y_0) + q_0|^2.$$

For this to hold with small $\alpha < 0$, it becomes necessary that the coefficient of α is zero. That is to say, $\lambda(x_0 - y_0) + q_0 = 0$, and rearranging this expression tells us

$$y_0 = x_0 + \frac{q_0}{\lambda}.$$

Combining this with 3.13 gives us

$$(q_0, Y_0) \in J^{2,+} f(x_0 + \frac{x_0}{\lambda}),$$

and making the substitution in 3.11 yields

$$f_\lambda(x_0) + \frac{1}{2\lambda}|q_0|^2 = f(x_0 + \frac{q_0}{\lambda}).$$

This is exactly what we wanted to show, so we are done, modulo the claim from the very beginning of this proof. But this is just a calculation:

$$\begin{aligned} f(x) - \frac{\lambda}{2}|y-x|^2 &\leq f_\lambda(y) \\ &\leq f_\lambda(x_0) + \langle q_0, y - x_0\rangle \\ &\qquad + \langle Y(y-x_0), y - x_0\rangle + o(|y-x_0|^2) \\ &= f(y_0) - \frac{\lambda}{2}|y_0 - x_0|^2 \\ &\qquad + \langle q_0, y - x_0\rangle + \frac{1}{2}\langle Y(y-x_0), y-x_0\rangle + o(|y-x_0|^2). \end{aligned}$$

The first inequality came from the definition of sup convolution; the second followed from the assumption that $(q_0, Y_0) \in J^{2,+}f_\lambda(x_0)$; and the final equality results from our choice of y_0.

■

Corollary 3.6 *Under the assumptions of Lemma 3.5, if* $(0, Y) \in \overline{J}^{2,+}f_\lambda(0)$*, then also* $(0, Y) \in \overline{J}^{2,+}f(0)$.

There is nothing special here about the superjet being located at the point $x_0 = 0$; we just use this convention to simplify the notation throughout this and the later proofs.

Proof: If $(0, Y) \in \overline{J}^{2,+}f_\lambda(0)$, then there exist $x_j \in \mathbb{R}^n$ such that $x_j \to 0$ as $j \to \infty$, and there exist $(q_j, Y_j) \in \mathbb{R}^n \times S(\mathbb{R})$ so that

$$(q_j, Y_j) \in J^{2,+}f_\lambda(x_j) \quad \text{and} \quad (x_j, f_\lambda(x_j), q_j, Y_j) \to (0, 0, 0, Y) \ \text{ as } j \to \infty.$$

This is just the definition of the closure of a superjet. By Lemma 3.5, we obtain

$$(q_j, Y_j) \in J^{2,+}f(x_j + \frac{q_j}{\lambda}) \qquad \text{and} \qquad f(x_j + \frac{q_j}{\lambda}) = f_\lambda(x_j) + \frac{1}{2\lambda}|q_j|^2.$$

So far we have $(x_j + \frac{q_j}{\lambda}) \to 0$, $q_j \to 0$, and $Y_j \to 0$. If we can show that

$$f(x_j + \frac{q_j}{\lambda}) \to f(0) \qquad \text{as } j \to \infty,$$

then combining this with the fact $(q_j, Y_j) \in J^{2,+}f(x_j + \frac{q_j}{\lambda})$ will show that

$$(0, Y) \in J^{2,+}f(0).$$

Because f is upper semicontinuous, we have

$$\begin{aligned} f(0) &\geq \limsup_{j\to\infty} f(x_j + \frac{q_j}{\lambda}) \\ &= \limsup_{j\to\infty} \left(f_\lambda(x_j) + \frac{1}{2\lambda} \mid q_j \mid^2 \right) && \text{(by Lemma 3.5)} \\ &= f_\lambda(0) && \text{(since } f_\lambda \text{ is continuous)} \\ &\geq f(0), \end{aligned}$$

where in the last line we used the fact that $f(0) \leq f_\lambda(0)$, which follows directly from the definition of sup convolution.

That is to say, $f(0) = \limsup_{j\to\infty} f(x_j + \frac{q_j}{\lambda})$. Passing to a subsequence where the lim sup becomes an ordinary limit, we have thus finished the proof.

■

3.3 The Maximum Principle

We now come to the central theorem of this essay. The proof will make use of all the machinery we have built up thus far. This theorem gives us the nonempty semijets we need to complete the viscosity-solution analogue of the classical maximum principle discussed in Section 2.2.

Theorem 3.7 *Put* $\phi(x,y) = \frac{\alpha}{2} \mid x - y \mid^2$ *on* $\mathbb{R}^n \times \mathbb{R}^n$*, with* $\alpha > 0$*. Set*

$$w(x,y) = u(x) - v(y)$$

for $(x,y) \in \mathbf{\Omega} \times \mathbf{\Omega}$*, where* $u \in USC(\mathbf{\Omega})$ *and* $v \in LSC(\mathbf{\Omega})$*. Suppose* (x_0, y_0) *is a local maximum of* $w - \phi$ *relative to* $\mathbf{\Omega} \times \mathbf{\Omega}$*. Then there exist* $X, Y \in S(n)$ *such that*

$$(D_x\phi(x_0,y_0), X) = (\alpha(x_0 - y_0), X) \in \overline{J}^{2,+}_{\mathbf{\Omega}} u(x_0) \qquad \text{and} \tag{3.16}$$

$$(-D_y\phi(x_0,y_0), Y) = (\alpha(x_0 - y_0), Y) \in \overline{J}^{2,-}_{\mathbf{\Omega}} v(y_0). \tag{3.17}$$

Moreover,

$$\begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq 3\alpha \begin{pmatrix} I & -I \\ -I & I \end{pmatrix}$$

Note that the final matrix inequality implies $X \leq Y$, since the matrix on the right annihilates vectors of the form (ξ, ξ).

Proof: To simplify the notation later on, write $u_1 = u$ and $u_2 = -v$ and

$$A = D^2\phi = \alpha \begin{pmatrix} I & -I \\ -I & I \end{pmatrix}$$

.

Step 1: We first prove the result for the case $\mathbf{\Omega} = \mathbb{R}^n$ and $(x_0, y_0) = 0$, $u_i(0) = 0$, and each u_i is bounded above. We will reduce the general case to this one at the end of the proof. We now have

$$w(x,y) = u_1(x) + u_2(y) \leq \frac{1}{2}\langle A(x,y), (x,y)\rangle. \tag{3.18}$$

Here, (x, y) is represents the vector $\eta \in \mathbb{R}^{2n}$ defined by $\eta_i = x_i$ for $i = 1, ..., n$ and $\eta_i = y_{i-2n}$ for $i = n+1, ..., 2n$. Choose $\epsilon > 0$ and put

$$\lambda = \frac{1}{\epsilon} + \| A \|, \tag{3.19}$$

where the matrix norm is the supremum of the magnitudes of the eigenvalues of A. We claim that

$$\langle A\eta, \eta \rangle \leq \langle (A + \epsilon A^2)\xi, \xi \rangle + \left(\frac{1}{\epsilon} + \| A \| \right) |\eta - \xi|^2,$$

for all $\eta, \xi \in \mathbb{R}^n$. This claim will be dealt with at the end of this proof. Using this claim allows us to write

$$\left(u_1(x) - \frac{\lambda}{2} \mid x - \xi_1 \mid^2 \right) + \left(u_2(y) - \frac{\lambda}{2} \mid y - \xi_2 \mid^2 \right) \leq \frac{1}{2} \langle (A + \epsilon A^2)\xi, \xi \rangle \tag{3.20}$$

for $\xi = (\xi_1, \xi_2) \in \mathbb{R}^{2n}$. Hence

$$w(x, y) - \frac{\lambda}{2} \mid (x, y) - \xi \mid^2 \leq \frac{1}{2} \langle (A + \epsilon A^2)\xi, \xi \rangle. \tag{3.21}$$

Using Definition 3.3 and the value of λ just determined, we find that the sup convolutions satisfy an "additive separation of variables" property,

$$w_\lambda(\xi) = (u_1)_\lambda(\xi_1) + (u_2)_\lambda(\xi_2).$$

Observe that $(u_i)_\lambda(0) \geq u_i(0) = 0$, while $(u_1)_\lambda(0) + (u_2)_\lambda(0) \leq 0$. Therefore, $u_i(0) = 0$. Write $B = A + \epsilon A^2$ and apply Lemma 3.4 to w_λ. The matrix given by the lemma is block diagonal of the form

$$\begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix}$$

for some $X, Y \in S(n)$. We thus obtain $(0, X) \in \overline{J}^2(u_1)_\lambda(0)$ and $(0, -Y) \in \overline{J}^2(u_2)_\lambda(0)$. This lemma also tells us that

$$-\left(\frac{1}{\epsilon} + \| A \|\right) I \leq \begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq A + \epsilon A^2. \tag{3.22}$$

Replace A with what it represented, $D^2\phi(x_0, y_0)$, and choose $\epsilon = \frac{1}{\alpha}$ to obtain the desired matrix inequality

$$\begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq 3\alpha \begin{pmatrix} I & -I \\ -I & I \end{pmatrix}. \tag{3.23}$$

Finally, apply Lemma 3.5 to each u_i to conclude that $(0, X) \in \overline{J}^{2,+}u_1(0)$ and $(0, -Y) \in \overline{J}^{2,+}u_2(0)$. That is,

$$(0, X) \in \overline{J}^{2,+}u(0) \qquad \text{and} \qquad (0, Y) \in \overline{J}^{2,-}v(0).$$

Step 2: Now we show how to reduce the general case to the specific one already proved.

Let $\mathbf{\Omega}$ be an open subset of $\mathbb{R}^n$. Continuing to write $u_1 = u$ and $u_2 = -v$, we restrict u_1 to a compact set K_1 containing x_0 as an interior point. Similarly, restrict u_2 to a compact set K_2 containing y_0 as an interior point. Furthermore, we require K_1 and K_2 to be small enough that (x_0, y_0) is the absolute maximum of $w - \phi$ on $K_1 \times K_2$. Now extend these restrictions to $\mathbb{R}^n$ by (abusing notation and) setting

$$u_i(x_i) = \begin{cases} u_i(x_i) & if\ x_i \in K_i \\ -\infty & if\ x_i \in \mathbb{R}^n - K_i \end{cases}.$$

Because the K_i are compact, we have $u_i \in USC(\mathbb{R}^n)$. Because x_0 is an interior point of $K_1 \subset \mathbf{\Omega}$, we have

$$\overline{J}^{2,+}_{\mathbb{R}^n} u_1(x_0) = \overline{J}^{2,+}_{\mathbf{\Omega}} u_1(x_0) \tag{3.24}$$

by the remarks in Section 2.5. Similarly,

$$\overline{J}^{2,+}_{\mathbb{R}^n} u_2(y_0) = \overline{J}^{2,+}_{\mathbf{\Omega}} u_2(y_0).$$

By setting the extension equal to $-\infty$, we have guaranteed that (x_0, y_0) is still a local maximum of $w - \phi$, now relative to $\mathbb{R}^{2n}$. Next write $\tilde{u}_1(x) = u_1(x + x_0) - u_1(x_0)$, and $\tilde{u}_2(y) = u_2(y + y_0) - u_2(y_0)$. These definitions translate u_1 and u_2 to bring x_0 and y_0, respectively, to the origin, then they make 0 the (absolute maximum) values of u_1 and u_2 there. Since we have only translated the u_i in directions orthogonal to the coordinate axes, clearly

$$\overline{J}^{2,+}_{\mathbb{R}^n} \tilde{u}_1(0) = \overline{J}^{2,+}_{\mathbb{R}^n} u_1(x_0),$$

and combining this with equation 3.24 yields

$$\overline{J}^{2,+}_{\mathbf{\Omega}} u_1(0) = \overline{J}^{2,+}_{\mathbb{R}^n} \tilde{u}_1(x_0).$$

Similar equalities hold for the second order superjet of u_2 at $\tilde{y}$. Also note that the origin is a local maximum of $\tilde{u}_1 + \tilde{u}_2 - \phi$. Applying the proof in Step 1 to the $\tilde{u}_i$ shows they have non-empty superjet closures at the origin, so u_1 and u_2 have non-empty superjet closures at x_0 and y_0, respectively. This completes the proof for the general case.

We still have to justify the claim that if $\epsilon > 0$ and $A \in S(m)$, then

$$\langle A\eta, \eta \rangle \leq \langle (A + \epsilon A^2)\xi, \xi \rangle + \left(\frac{1}{\epsilon} + \parallel A \parallel \right) \mid \eta - \xi \mid^2, \tag{3.25}$$

for $\eta, \xi \in \mathbb{R}^m$. To prove this, we will make use of the Cauchy-Schwarz inequality, the Cauchy inequality with epsilon and the fact that A is self-adjoint, so that we can write $|A\xi|^2 = \langle A\xi, A\xi\rangle = \langle A^2\xi, \xi\rangle$. These give us

$$\begin{aligned}
\langle A\eta, \eta\rangle &= \langle A(\xi + \eta - \xi), \xi + \eta - \xi\rangle \\
&= \langle A\xi, \xi\rangle + \langle A(\eta - \xi), \xi\rangle + \langle A\xi, \eta - \xi\rangle + \langle A(\eta - \xi), \eta - \xi\rangle \\
&\leq \langle A\xi, \xi\rangle + \| A \| | \eta - \xi |^2 + 2 \, | \langle \eta - \xi, A\xi\rangle | \\
&\leq \langle A\xi, \xi\rangle + \| A \| | \eta - \xi |^2 + 2 \, (|\eta - \xi||A\xi|) \\
&\leq \langle A\xi, \xi\rangle + \| A \| | \eta - \xi |^2 + 2 \left(\frac{1}{2\epsilon} \, | \eta - \xi |^2 + \frac{\epsilon}{2}\langle A^2\xi, \xi\rangle \right) \\
&= \langle (A + \epsilon A^2)\xi, \xi\rangle + \left(\frac{1}{\epsilon} + \| A \| \right) | \eta - \xi |^2 .
\end{aligned}$$

■

3.4 The Comparison Principle

We now wish to extend the comparison principle discussed in Section 2.2 to the case $u \in USC(\mathbf{\Omega})$ and $v \in LSC(\mathbf{\Omega})$. Here, u is a subsolution and v is a supersolution of some second order PDE

$$F(x, w(x), Dw(x), D^2w(x)) = 0,$$

with F proper. We want to show that if $u \leq v$ on $\partial\mathbf{\Omega}$, then $u \leq v$ on $\mathbf{\Omega}$. If u and v were twice differentiable (i.e. they are a classical subsolution and supersolution, respectively), then the argument would be simple, if we make the additional assumption that F is strictly increasing in the second variable:

At any interior local maximum $\tilde{x}$ of $u - v$,

$$Du(\tilde{x}) = Dv(\tilde{x}) \qquad \text{and} \qquad D^2u(\tilde{x}) \leq D^2v(\tilde{x}).$$

Hence

$$F(\tilde{x}, u(\tilde{x}), Du(\tilde{x}), D^2u(\tilde{x})) \leq 0$$

$$\leq F(\tilde{x}), v(\tilde{x}), Dv(\tilde{x}), D^2v(\tilde{x}))$$

$$\leq F(\tilde{x}), v(\tilde{x}), Du(\tilde{x}), D^2u(\tilde{x})).$$

The first inequality follows since u is a subsolution, the second because v is a supersolution, and the third because F is proper. Now, by the strict monotonicity assumption on F, we conclude that $u(\tilde{x}) \leq v(\tilde{x})$. Since this is true at any interior local maximum of $u - v$, and $u \leq v$ on $\partial\boldsymbol{\Omega}$, we must have $u \leq v$ in $\boldsymbol{\Omega}$.

Now if u and v are not classical solutions, the proof of this comparison principle becomes much more intricate. We do not have $Du(\tilde{x})$, and $D^2u(\tilde{x})$ to plug into F, nor do we have derivatives for v. We want to use the semijets $J^{2,+}u$ and $J^{2,-}v$, but these may be empty, even at interior local maxima of $u - v$.

To compensate for this fact, we approximate the interior local maxima of $u - v$ using the following lemma.

Lemma 3.8 *Let $\boldsymbol{\Omega} \subset \mathbb{R}^n$ be a bounded domain and $u \in USC(\overline{\boldsymbol{\Omega}})$ and $v \in LSC(\overline{\boldsymbol{\Omega}})$ be real-valued. Define*

$$M_\alpha = \sup_{\overline{\boldsymbol{\Omega}} \times \overline{\boldsymbol{\Omega}}} \left(u(x) - v(y) - \frac{\alpha}{2} \mid x - y \mid^2 \right),$$

for $\alpha > 0$. Suppose (x_α, y_α) are chosen such that

$$\lim_{\alpha \to \infty} \left(M_\alpha - \left(u(x_\alpha) - v(y_\alpha) - \frac{\alpha}{2} \mid x_\alpha - y_\alpha \mid^2 \right) \right) = 0. \tag{3.26}$$

Then the following are true:

$$(i) \quad \lim_{\alpha \to \infty} \alpha \mid x_\alpha - y_\alpha \mid^2 = 0 \qquad \textit{and} \tag{3.27}$$

$$(ii) \quad \lim_{\alpha \to \infty} M_\alpha = u(\hat{x}) - v(\hat{x}) = \sup_{\overline{\boldsymbol{\Omega}}} (u(x) - v(x)), \tag{3.28}$$

whenever $\hat{x}$ is a limit point of x_α as $\alpha \to \infty$.

Proof:

We introduce $\delta_\alpha = M_\alpha - \left(u(x_\alpha) - v(y_\alpha) - \frac{\alpha}{2}|x_\alpha - y_\alpha|^2\right)$. Because of the upper-semicontinuity and the fact that $\overline{\mathbf{\Omega}} \times \overline{\mathbf{\Omega}}$ is compact, we see that $M_\alpha < \infty$, and by the assumption 3.26 it is clear that $\delta_\alpha \to 0$ as $\alpha \to \infty$.

Since $|x_\alpha - y_\alpha|^2 \geq 0$, for each fixed $x, y \in \overline{\mathbf{\Omega}}$ we see that $u(x) - v(y) - \frac{\alpha}{2}|x-y|^2$ decreases as α increases. Consequently the supremum over all $x, y \in \overline{\mathbf{\Omega}}$ decreases, and M_α decreases as α increases.

Note that

$$M_\alpha \geq u(x) - v(x)$$

for any $x \in \overline{\mathbf{\Omega}}$. Hence $\lim_{\alpha\to\infty} M_\alpha$ exists and is finite. Also observe that

$$M_{\alpha/2} \geq u(x_\alpha) - v(y_\alpha) - \frac{\alpha}{4}|x_\alpha - y_\alpha|^2 \geq M_\alpha - \delta_\alpha + \frac{\alpha}{4}|x_\alpha - y_\alpha|^2.$$

Thus $\frac{\alpha}{2}|x_\alpha - y_\alpha|^2 \leq 2(M_{\alpha/2} - M_\alpha + \delta_\alpha)$. The right side of this inequality goes to zero as $\alpha \to \infty$. Therefore,

$$\frac{\alpha}{2}|x_\alpha - y_\alpha|^2 \to 0 \qquad \text{as } \alpha \to \infty.$$

This proves 3.27. As an obvious weaker consequence, $|x_\alpha - y_\alpha| \to 0$. Suppose now that $(\hat{x}, \hat{y}) \in \overline{\mathbf{\Omega}} \times \overline{\mathbf{\Omega}}$ is a limit point of $\{(x_\alpha, y_\alpha)\}$ as $\alpha \to \infty$. Then there exists a sequence $\alpha_n \to \infty$ such that $x_{\alpha_n} \to \hat{x}$ and $y_{\alpha_n} \to \hat{y}$ as $n \to \infty$, thus

$$|\hat{x} - \hat{y}| = \lim_{n\to\infty} |x_\alpha - y_\alpha| = 0.$$

That is to say, $\hat{x} = \hat{y}$. Observe that

$$u(x_{\alpha_n}) - v(y_{\alpha_n}) = M_{\alpha_n} - \delta_{\alpha_n} \geq \sup_{z\in\overline{\mathbf{\Omega}}} \left(u(z) - v(z)\right) - \delta_{\alpha_n}.$$

Because $\Phi(x, y) = u(x) - v(y)$ is upper semicontinuous, $\delta_{\alpha_n} \to 0$ and $\hat{x} = \hat{y}$, we thus have

$$\sup_{z\in\overline{\mathbf{\Omega}}} \left(u(z) - v(z)\right) \geq u(\hat{x}) - v(\hat{y}) \geq \lim_{\alpha\to 0} M_\alpha \geq \sup_{z\in\overline{\mathbf{\Omega}}} \left(u(z) - v(z)\right).$$

That is,

$$\lim_{\alpha\to 0} M_\alpha = \sup_{z\in\overline{\mathbf{\Omega}}} \left(u(z) - v(z)\right).$$

This proves 3.28.

■

We are finally ready to prove a viscosity-solution analogue of the classical comparison principle.

Theorem 3.9 (The Comparison Principle) *Let $\mathbf{\Omega}$ be a bounded open subset of $\mathbb{R}^n$, $F \in C(\mathbf{\Omega} \times \mathbb{R} \times \mathbb{R}^n \times S(n))$ be proper and satisfy*

(i) There exists $\gamma > 0$ such that

$$\gamma(r - s) \le F(x, r, p, X) - F(x, s, p, X)$$

for $r \ge s$, $(x, p, X) \in \overline{\mathbf{\Omega}} \times \mathbb{R}^n \times S(n)$ and

(ii) $F(x, r, p, X)$ is uniformly continuous in the spatial variable x.

Suppose $u \in USC(\overline{\mathbf{\Omega}})$ is a subsolution of $F = 0$ in $\mathbf{\Omega}$, and $v \in LSC(\overline{\mathbf{\Omega}})$ is a supersolution of $F = 0$ in $\mathbf{\Omega}$.

If $u \le v$ on $\partial\mathbf{\Omega}$, then $u \le v$ in $\overline{\mathbf{\Omega}}$.

Proof:

Suppose, to get a contradiction, that $u > v$ at some interior point of $\mathbf{\Omega}$. For each $m \in \mathbb{N}$, we can find $x_m, y_m \in \overline{\mathbf{\Omega}} \times \overline{\mathbf{\Omega}}$ such that

$$\left(u(x_m) - v(y_m) - \frac{m}{2}|x_m - y_m|^2\right) = \sup_{\overline{\mathbf{\Omega}}\times\overline{\mathbf{\Omega}}} \left(u(x) - v(y) - \frac{m}{2}|x - y|^2\right).$$

(These points can be found in view of upper-semicontinuity and compactness.) The sequence (x_m, y_m) has a limit point in the compact set $\overline{\mathbf{\Omega}} \times \overline{\mathbf{\Omega}}$. Pass to a subsequence so that we can write $\lim_{m\to\infty}(x_m, y_m) = (\hat{x}, \hat{y})$. By Lemma 3.8, $m|x_m - y_m|^2 \to 0$, so $\hat{x} = \hat{y}$. Let us write z for this point, so that $\lim_{m\to\infty}(x_m, y_m) = (z, z)$. By Lemma 3.8 again, we have $u(z) - v(z) = \sup_{\overline{\mathbf{\Omega}}}(u(x) - v(x))$. We have assumed that $u \le v$ on $\partial\mathbf{\Omega}$ but $u > v$ somewhere in $\mathbf{\Omega}$. Hence $z \notin \partial\mathbf{\Omega}$, leaving us with $z \in \mathbf{\Omega}$.

Since $(x_m, y_m) \to (z, z)$, we have $(x_m, y_m) \in \mathbf{\Omega} \times \mathbf{\Omega}$ for large enough m (they may be on the boundary for small m). Restricting our attention to large m from now on, we see that (x_m, y_m) is a local maximum of $u(x) - v(y) - \frac{m}{2}|x-y|^2$ relative to $\mathbf{\Omega} \times \mathbf{\Omega}$. By Theorem 3.7, there exist matrices $X_m, Y_m \in S(n)$ such that

$$(m(x_m - y_m), X_m) \in \overline{J}_{\mathbf{\Omega}}^{2,+} u(x_m) \qquad \text{and} \qquad (m(x_m - y_m), Y_m) \in \overline{J}_{\mathbf{\Omega}}^{2,-} v(y_m).$$

Because u is a subsolution and v is a supersolution, we have

$$F(x_m, u(x_m), m(x_m - y_m(, X_m) \leq 0 \leq F(y_m, v(y_m), m(x_m - y_m), Y_m)$$

(see the remarks at the end of section 2.4). Using our coercitivity hypothesis on the second variable of F and the assumption that there exists $\delta > 0$ such that

$$\delta \leq M_m = u(x_m) - v(y_m) - \frac{m}{2}|x_m - y_m|^2 \leq u(x_m) - v(y_m),$$

we can write

$$\begin{aligned}
0 &< \gamma\delta \\
&\leq \gamma(u(x_m) - v(y_m)) \\
&\leq F(x_m, u(x_m), m(x_m - y_m), X_m) - F(x_m, v(y_m), m(x_m - y_m), X_m) \\
&\leq F(x_m, u(x_m), m(x_m - y_m), X_m) - F(y_m, v(y_m), m(x_m - y_m), Y_m) \\
&\quad F(y_m, v(y_m), m(x_m - y_m), Y_m) - F(x_m, v(y_m), m(x_m - y_m), X_m).
\end{aligned}$$

On the right side of this equation, the difference in the first pair of terms is majorized by zero because u is a subsolution and v is a supersolution. The difference in the second pair of terms can be analyzed similarly:

$$\begin{aligned}F(y_m,&v(y_m), m(x_m - y_m), Y_m) - F(x_m, v(y_m), m(x_m - y_m), X_m)\\&= F(y_m, v(y_m), m(x_m - y_m), Y_m) - F(y_m, v(y_m), m(x_m - y_m), X_m)\\&\quad + F(y_m, v(y_m), m(x_m - y_m), X_m) - F(x_m, v(y_m), m(x_m - y_m), X_m).\end{aligned}$$

The difference in the first pair of terms on the right side is again majorized by zero, this time because they differ only in the last variable. The function F is assumed to be proper, so it is non-increasing in that variable, while we know that $X \leq Y$.

We now have

$$\begin{aligned}0 &< \gamma\delta\\&\leq F(y_m, v(y_m), m(x_m - y_m), X_m) - F(x_m, v(y_m), m(x_m - y_m), X_m)\\&\leq \omega(|x_m - y_m|),\end{aligned}$$

where we have applied our hypothesized uniform continuity condition on the first variable of F by introducing this function ω. Now letting $m \to \infty$, $|x_m - y_m| \to 0$ by Lemma 3.8, and

$$\lim_{m\to\infty} \omega(|x_m - y_m|) = 0.$$

This gives us $0 < \gamma\delta \leq 0$, the desired contradiction.

■

As a final note, we give one useful application of the comparison principle. Suppose $\mathbf{\Omega}$ and F satisfy the hypotheses of Theorem 3.9, and u and v are both viscosity solutions of the Dirichlet problem

$$\begin{cases} F(x, w(x), Dw(x), D^2w(x)) = 0 & \text{in } \mathbf{\Omega} \\ \qquad\qquad \text{and} \quad w(x) = f(x) & \text{on } \partial\mathbf{\Omega} \end{cases},$$

for some given boundary values $f(x)$. Then $u \leq v$ on $\partial\mathbf{\Omega}$, u is a subsolution and v is a supersolution, so $u \leq v$ in $\mathbf{\Omega}$. Similarly, $v \leq u$ on $\partial\mathbf{\Omega}$, v is a subsolution and u is a supersolution, so $v \leq u$ in $\mathbf{\Omega}$.

That is to say, given certain coercitivity and continuity conditions on the function F, viscosity solutions of the Dirichlet problem are unique.

Bibliography

Bibliography

[CIL] Crandall, Michael, Ishii, Hitoshi, and Lions, Pierre-Louis, User's Guide to Viscosity Solutions of Second Order Partial Differential Equations, Bull. of Amer. Math. Soc., Vol. 27, No. 1 (1992), pp. 1–67.

[E] Evans, Lawrence C., *Partial Differential Equations,* Graduate Studies in Mathematics, Vol. 19, Amer. Math. Soc., 1998.

[EG] Evans, L. C. and Gariepy, R., *Measure Theory and Fine Properties of Functions,* Studies in Advanced Mathematics, CRC Press, 1992.

[J] Jensen, R., The Maximum Principle for Viscosity Solutions of Fully Nonlinear Second Order Partial Differential Equations, Arch. Rat. Mech. Anal., Vol. 101, (1988), pp. 1–27

[R] Rudin, Walter, *Real and Complex Analysis,* (3d ed.), McGraw-Hill, 1987.