Economics 204

Lecture 11–Monday, August 10, 2009

Sections 4.1-4.3, Unified Treatment

Definition 1 Let $f : I \to \mathbf{R}$, where $I \subseteq \mathbf{R}$ is an open interval. f is *differentiable* at $x \in I$ if

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = a$$

for some $a \in \mathbf{R}$.

This is equivalent to

$$\begin{aligned}
& \lim_{h \to 0} \frac{f(x+h) - (f(x) + ah)}{h} = 0 \\
\Leftrightarrow \quad & \forall_{\varepsilon > 0} \, \exists_{\delta > 0} \; 0 < |h| < \delta \Rightarrow \left| \frac{f(x+h) - (f(x) + ah)}{h} \right| < \varepsilon \\
\Leftrightarrow \quad & \forall_{\varepsilon > 0} \, \exists_{\delta > 0} \; 0 < |h| < \delta \Rightarrow \frac{|f(x+h) - (f(x) + ah)|}{|h|} < \varepsilon \\
\Leftrightarrow \quad & \lim_{h \to 0} \frac{|f(x+h) - (f(x) + ah)|}{|h|} = 0
\end{aligned}$$

Recall that the limit considers h near zero, but not $h = 0$.

Definition 2 If $X \subseteq \mathbf{R}^n$ is open, $f : X \to \mathbf{R}^m$ is *differentiable at* $x \in X$ if

$$\exists_{T_x \in L(\mathbf{R}^n, \mathbf{R}^m)} \lim_{h \to 0, h \in \mathbf{R}^n} \frac{|f(x+h) - (f(x) + T_x(h))|}{|h|} = 0$$

(Recall $| \cdot |$ denotes the Euclidean distance.) f is *differentiable* if it is differentiable at all $x \in X$.

h is a small, nonzero element of $\mathbf{R}^n$; $h \to 0$ from any direction, along a spiral, etc. One linear operator T_x works no matter how h approaches zero.

$$f(x) + T_x(h) \text{ is the best linear}$$
$$\text{approximation to } f(x+h) \text{ for small } h$$

Notation:

$$\begin{array}{rl} & y = O(|h|^n) \text{ as } h \to 0 \\ \text{means} & \exists_{K,\delta>0}\ |h| < \delta \Rightarrow |y| \le K|h|^n \\ \text{read} & y \text{ is big-Oh of } |h|^n \\ & y = o(|h|^n) \text{ as } h \to 0 \\ \text{means} & \lim_{h\to 0} \frac{|y|}{|h|^n} = 0 \\ \text{read} & y \text{ is little-oh of } |h|^n \end{array}$$

Note that the statement $y = O(|h|^{n+1})$ as $h \to 0$ implies $y = o(|h|^n)$ as $h \to 0$. Note that

$$f \text{ is differentiable at } x \Leftrightarrow \exists_{T_x \in L(\mathbf{R}^n, \mathbf{R}^m)}$$
$$f(x+h) = f(x) + T_x(h) + o(h) \text{ as } h \to 0$$

Notation:

$$\begin{array}{rcl} df_x & & \text{is the linear transformation } T_x \\ Df(x) & & \text{is the matrix of } df_x \text{ with respect to the} \\ & & \text{standard basis; called the } \textit{Jacobian} \\ & & \text{or } \textit{Jacobian matrix} \text{ of } f \text{ at } x \\ E_f(h) & = & f(x+h) - (f(x) + df_x(h)) \text{ (Error Term)} \\ & & f \text{ is differentiable at } x \\ & \Leftrightarrow & E_f(h) = o(h) \text{ as } h \to 0 \end{array}$$

Let's compute $Df(x) = (a_{ij})$. Let $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbf{R}^n$. Look in direction e_j; $|\gamma e_j| = |\gamma|$.

$$
\begin{aligned}
o(\gamma) & \\
&= f(x+\gamma e_j) - (f(x) + T_x(\gamma e_j)) \\
&= f(x+\gamma e_j) - \left(f(x) + \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mj} & \vdots & a_{mn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \gamma \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) \\
&= f(x+\gamma e_j) - \left(f(x) + \begin{pmatrix} \gamma a_{1j} \\ \vdots \\ \gamma a_{mj} \end{pmatrix} \right)
\end{aligned}
$$

For $i = 1, \ldots, m$, let f^i denote the i^{th} component of the function f:

$$
\begin{aligned}
f^i(x+\gamma e_j) - \left(f^i(x) + \gamma a_{ij} \right) &= o(\gamma) \\
\text{so } a_{ij} &= \frac{\partial f^i}{\partial x_j}
\end{aligned}
$$

Theorem 3 (3.3) *Suppose $X \subseteq \mathbf{R}^n$ is open and $f : X \to \mathbf{R}^m$ is differentiable at $x \in X$. Then $\frac{\partial f^i}{\partial x_j}$ exists for $1 \leq i \leq m$, $1 \leq j \leq n$, and*

$$
(Df)(x) = \begin{pmatrix} \frac{\partial f^1}{\partial x_1} & \cdots & \frac{\partial f^1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f^m}{\partial x_1} & \cdots & \frac{\partial f^m}{\partial x_n} \end{pmatrix}
$$

i.e. the Jacobian is the matrix of partial derivatives.

Remark: If f is differentiable at x, then all first-order partial derivatives $\frac{\partial f^i}{\partial x_j}$ exist at x. However, existence of all the first-order partial derivatives does not imply that f is differentiable.

Theorem 4 (3.4) *If all the first-order partial derivatives $\frac{\partial f^i}{\partial x_j}$ $(1 \leq i \leq m,\ 1 \leq j \leq n)$ exist and are continuous at x, then f is differentiable at x.*

Directional Derivatives: Suppose $X \subseteq \mathbf{R}^n$ open, $f : X \to \mathbf{R}^m$ differentiable at x, $|u| = 1$.

$$f(x+\gamma u) - (f(x) + T_x(\gamma u)) = o(\gamma) \text{ as } \gamma \to 0$$

$$\Rightarrow \quad f(x+\gamma u) - (f(x) + \gamma T_x(u)) = o(\gamma) \text{ as } \gamma \to 0$$

$$\Rightarrow \quad \lim_{\gamma \to 0} \frac{f(x+\gamma u) - f(x)}{\gamma} = T_x(u) = Df(x)u$$

i.e. the directional derivative in the direction u (with $|u| = 1$) is

$$Df(x)_{m \times n} u_{n \times 1} \in \mathbf{R}^m$$

Theorem 5 (3.5, Chain Rule) *Let $X \subseteq \mathbf{R}^n$, $Y \subseteq \mathbf{R}^m$ be open, $f : X \to Y$, $g : Y \to \mathbf{R}^p$. Let $x_0 \in X$, $F = g \circ f$. If f is differentiable at x_0 and g is differentiable at $f(x_0)$, then $F = g \circ f$ is differentiable at x_0 and*

$$\begin{aligned} dF_{x_0} &= dg_{f(x_0)} \circ df_{x_0} \\ &\quad \textit{(composition of linear transformations)} \\ DF(x_0) &= Dg(f(x_0))Df(x_0) \\ &\quad \textit{(matrix multiplication)} \end{aligned}$$

Remark: The statement is exactly the same as in the univariate case, except we replace the univariate derivative by a linear transformation. The proof is more or less the same, with a bit of linear algebra added.

Theorem 6 (1.7, Mean Value Theorem, Univariate Case) *Suppose $f : [a, b] \to \mathbf{R}$ is continuous on $[a, b]$ and differentiable on (a, b). Then there exists $c \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{b - a} = f'(c)$$

In other words,

$$f(b) - f(a) = f'(c)(b - a)$$

Remark: The Mean Value Theorem is useful for estimating bounds on functions and error terms in approximation of functions.

Proof: Consider the function

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a)$$

Then $g(a) = 0 = g(b)$. Note that for $x \in (a, b)$,

$$g'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}$$

so it suffices to find $c \in (a, b)$ such that $g'(c) = 0$.

Case I: If $g(x) = 0$ for all $x \in [a, b]$, choose an arbitrary $c \in (a, b)$, and note that $g'(c) = 0$, so we are done.

Case II: Suppose $g(x) > 0$ for some $x \in [a, b]$. Since g is continuous on $[a, b]$, it attains its maximum at some point $c \in (a, b)$. Since g is differentiable at c and c is an interior point of the domain of g, we have $g'(c) = 0$, and we are done.

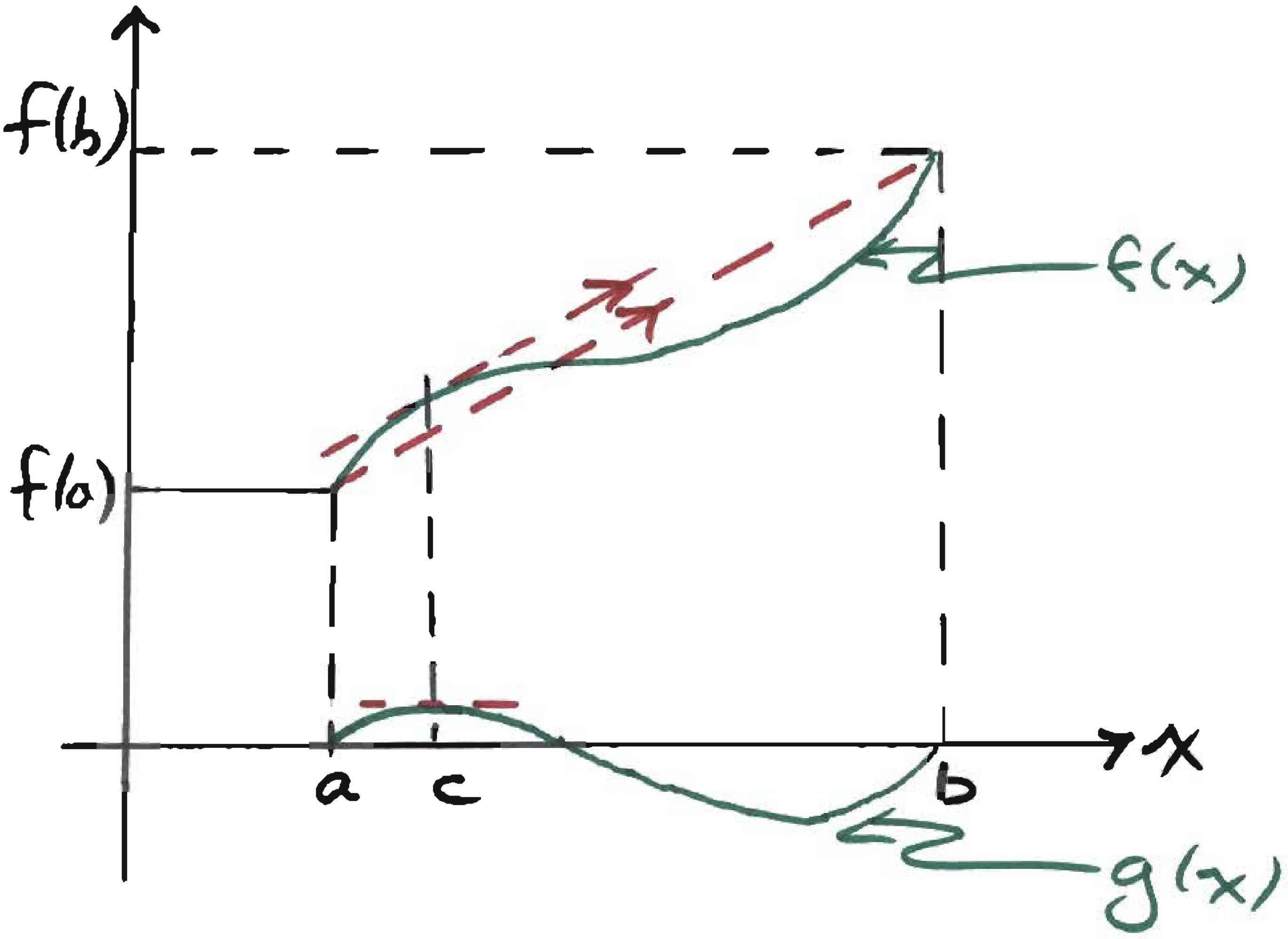
Case III: If $g(x) < 0$ for some $x \in [a, b]$, the argument is similar to that in Case II. ■

Notation:

$$L(x, y) = \{\alpha x + (1 - \alpha)y : \alpha \in [0, 1]\}$$

is the line segment from x to y.

Mean Value Theorem



Theorem 7 (Mean Value Theorem) *Suppose* $f : \mathbf{R}^n \to \mathbf{R}$ *is differentiable on an open set* $X \subseteq \mathbf{R}^n$*,* $x, y \in X$*,* $L(x, y) \subseteq X$*. Then there exists* $z \in L(x, y)$ *such that*

$$f(y) - f(x) = Df(z)(y - x)$$

Remark: This statement is different from Theorem 3.7 in de la Fuente. Notice that the statement is exactly the same as in the univariate case. For $f : \mathbf{R}^n \to \mathbf{R}^m$, we can apply the Mean Value Theorem to each component, to obtain $z_1, \ldots, z_m \in L(x, y)$ such that

$$f^i(y) - f^i(x) = Df^i(z_i)(y - x)$$

However, we cannot find a single z which works for every component.

The following result plays the same role in estimating function values and error terms for functions taking values in $\mathbf{R}^m$ as the Mean Value Theorem plays for functions from $\mathbf{R}$ to $\mathbf{R}$.

Theorem 8 *Suppose* $X \subset \mathbf{R}^n$ *is open,* $f : X \to \mathbf{R}^m$ *is differentiable. If* $x, y \in X$ *and* $L(x, y) \subseteq X$*, then there exists* $z \in L(x, y)$ *such that*

$$\begin{aligned} |f(y) - f(x)| &\leq |df_z(y - x)| \\ &\leq \|df_z\| |y - x| \end{aligned}$$

Remark: To understand why we don't get equality, consider $f : [0, 1] \to \mathbf{R}^2$ defined by

$$f(t) = (\cos 2\pi t, \sin 2\pi t)$$

f maps $[0, 1]$ to the unit circle in $\mathbf{R}^2$. Note that $f(0) = f(1) = (1, 0)$, so $|f(1) - f(0)| = 0$. However, for any $z \in [0, 1]$,

$$\begin{aligned} |df_z(1 - 0)| &= |2\pi(-\sin 2\pi t, \cos 2\pi t)| \\ &= 2\pi\sqrt{\sin^2 2\pi t + \cos^2 2\pi t} \\ &= 2\pi \end{aligned}$$

Section 4.4, Taylor's Theorem

Theorem 9 (1.9, Taylor's Theorem in $\mathbf{R}^1$) *Let $f : I \to \mathbf{R}$ be n-times differentiable, where $I \subseteq \mathbf{R}$ is an open interval. If $x, x+h \in I$, then*

$$
\begin{aligned}
f(x+h) &= f(x) + \sum_{k=1}^{n-1} \frac{f^{(k)}(x)h^k}{k!} + E_n \\
&\textit{where} \\
f^{(k)} &\textit{ is the } k^{th} \textit{ derivative of } f \\
E_n &= \frac{f^{(n)}(x+\lambda h)h^n}{n!} \textit{ for some } \lambda \in (0,1)
\end{aligned}
$$

Motivation: Let

$$
\begin{aligned}
T_n(h) &= f(x) + \sum_{k=1}^{n} \frac{f^{(k)}(x)h^k}{k!} \\
&= f(x) + f'(x)h + \frac{f''(x)h^2}{2} + \cdots + \frac{f^{(n)}(x)h^n}{n!} \\
T_n(0) &= f(x) \\
T_n'(h) &= f'(x) + f''(x)h + \cdots + \frac{f^{(n)}(x)h^{n-1}}{(n-1)!} \\
T_n'(0) &= f'(x) \\
T_n''(h) &= f''(x) + \cdots + \frac{f^{(n)}(x)h^{n-2}}{(n-2)!} \\
T_n''(0) &= f''(x) \\
&\vdots \\
T_n^{(n)}(0) &= f^{(n)}(x)
\end{aligned}
$$

so $T_n(h)$ is the unique n^{th} degree polynomial such that

$$\begin{aligned} T_n(0) &= f(x) \\ T_n'(0) &= f'(x) \\ &\vdots \\ T_n^{(n)}(0) &= f^{(n)}(x) \end{aligned}$$

The proof of the formula for the remainder E_n is essentially the Mean Value Theorem; the problem in applying it is that the point $x + \lambda h$ is not known in advance.

Theorem 10 (Alternate Taylor's Theorem in $\mathbf{R}^1$) *Let $f : I \to \mathbf{R}$ be n times differentiable, where $I \subseteq \mathbf{R}$ is an open interval and $x \in I$. Then*

$$f(x+h) = f(x) + \sum_{k=1}^{n} \frac{f^{(k)}(x)h^k}{k!} + o\left(h^n\right) \text{ as } h \to 0$$

If f is $(n+1)$ times continuously differentiable (i.e. all derivatives up to order $n+1$ exist and are continuous), then

$$f(x+h) = f(x) + \sum_{k=1}^{n} \frac{f^{(k)}(x)h^k}{k!} + O\left(h^{n+1}\right) \text{ as } h \to 0$$

Remark: The first equation in the statement of the theorem is essentially a restatement of the definition of the n^{th} derivative. The second statement is proven from Theorem 1.9, and the continuity of the derivative, hence the boundedness of the derivative on a neighborhood of x.

Definition 11 $X \subseteq \mathbf{R}^n$, X open, $f : X \to \mathbf{R}^m$. f is *continuously differentiable* on X if

- f is differentiable on X and

- df_x is a continuous function of x from X to $L(\mathbf{R}^n, \mathbf{R}^m)$, with operator norm $\|df_x\|$

f is C^k if all partial derivatives of order $\leq k$ exist and are continuous in X.

Theorem 12 (4.3) *Suppose $X \subseteq \mathbf{R}^n$, X open, $f : X \to \mathbf{R}^m$. Then f is continuously differentiable on X if and only if f is C^1.*

Notational Problem in Taylor's Theorem: If $f : \mathbf{R}^n \to \mathbf{R}^m$, the quadratic terms are OK for $m = 1$; for $m > 1$, handle each component separately. For cubic and higher order terms, notation is a mess.

Linear Terms:

Theorem 13 *Suppose $X \subseteq \mathbf{R}^n$, X is open, $x \in X$. If $f : X \to \mathbf{R}^m$ is differentiable, then*

$$f(x + h) = f(x) + Df(x)h + o(h) \text{ as } h \to 0$$

The previous theorem is essentially a restatement of the definition of differentiability.

Theorem 14 (Corollary of 4.4) *Suppose $X \subseteq \mathbf{R}^n$, X is open, $x \in X$. If $f : X \to \mathbf{R}^m$ is C^2, then*

$$f(x + h) = f(x) + Df(x)h + O\left(|h|^2\right) \text{ as } h \to 0$$

Quadratic Terms:

Treat each component of the function separately, so consider $f : X \to \mathbf{R}$, $X \subseteq \mathbf{R}^n$ an open set.

Let

$$D^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

$$f \in C^2 \Rightarrow \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

$$\Rightarrow D^2 f(x) \text{ is symmetric}$$

$$\Rightarrow D^2 f(x) \text{ has an orthonormal basis of eigenvectors}$$

$$\text{and thus can be diagonalized}$$

Theorem 15 (Stronger Version of 4.4) *Let $X \subseteq \mathbf{R}^n$ be open, $f : X \to \mathbf{R}$, $f \in C^2(X)$, $x \in X$. Then*

$$\begin{aligned} f(x+h) &= f(x) + Df(x)h \\ &\quad + \frac{1}{2} h^\top (D^2 f(x))h + o\left(|h|^2\right) \text{ as } h \to 0 \end{aligned}$$

If $f \in C^3$,

$$\begin{aligned} f(x+h) &= f(x) + Df(x)h \\ &\quad + \frac{1}{2} h^\top (D^2 f(x))h + O\left(|h|^3\right) \text{ as } h \to 0 \end{aligned}$$

Remark: De la Fuente assumes X is convex which he has not yet defined. X is said to be *convex* if, for every $x, y \in X$ and $\alpha \in [0,1]$, $\alpha x + (1-\alpha)y \in X$. We don't need this. Since X is open,

$$x \in X \Rightarrow \exists_{\delta > 0} B_\delta(x) \subseteq X$$

and $B_\delta(x)$ is convex.

Definition 16 We say f has a *saddle* at x if $Df(x) = 0$ but x has neither a local maximum nor a local minimum at x.

Corollary 17 *Suppose $X \subseteq \mathbf{R}^n$, X is open, $x \in X$. If $f : X \to \mathbf{R}$ is C^2, there is an orthonormal basis $\{v_1, \ldots, v_n\}$ and corresponding eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbf{R}$ such that*

$$
\begin{aligned}
& f(x+h) \\
&= f(x + \gamma_1 v_1 + \cdots + \gamma_n v_n) \\
&= f(x) + \sum_{i=1}^{n} (Df(x)v_i)\,\gamma_i \\
&\quad + \frac{1}{2}\sum_{i=1}^{n} \lambda_i \gamma_i^2 + o\left(|\gamma|^2\right)
\end{aligned}
$$

where $\gamma_i = h \cdot v_i$.

- *If $f \in C^3$, we may strengthen $o(|\gamma|^2)$ to $O(|\gamma|^3)$.*
- *If f has a local maximum or local minimum at x, then*

$$Df(x) = 0$$

- *If $Df(x) = 0$, then*

$$
\begin{aligned}
& \lambda_1, \ldots, \lambda_n > 0 \\
& \quad \Rightarrow \quad f \text{ has a local minimum at } x \\
& \lambda_1, \ldots, \lambda_n < 0 \\
& \quad \Rightarrow \quad f \text{ has a local maximum at } x \\
& \lambda_i < 0 \text{ for some } i,\ \lambda_j > 0 \text{ for some } j \\
& \quad \Rightarrow \quad f \text{ has a saddle at } x \\
& \lambda_1, \ldots, \lambda_n \geq 0,\ \lambda_i > 0 \text{ for some } i \\
& \quad \Rightarrow \quad f \text{ has a local minimum} \\
& \qquad\quad \text{or a saddle at } x
\end{aligned}
$$

$$\lambda_1, \ldots, \lambda_n \leq 0,\ \lambda_i < 0 \textit{ for some } i$$

$$\Rightarrow \quad f \textit{ has a local maximum}$$

$$\textit{or a saddle at } x$$

$$\lambda_1 = \cdots = \lambda_n = 0 \textit{ gives no information.}$$

Proof: The idea is that the error term tells us that the local behavior is dominated by the quadratic terms. From our study of quadratic forms, we know the behavior of the quadratic terms is determined by the signs of the eigenvalues. If $\lambda_i = 0$ for some i, then we know that the quadratic form arising from the second partial derivatives is identically zero in the direction v_i, and the higher derivatives will determine the behavior of the function f in the direction v_i. For example, if $f(x) = x^3$, then $f'(0) = 0$, $f''(0) = 0$, but we know that f has a saddle at $x = 0$; however, if $f(x) = x^4$, then again $f'(0) = 0$ and $f''(0) = 0$ but f has a local (and global) minimum at $x = 0$.■