R.P. McAfee

October, 1986

ECONOMICS 241

HANDOUT 4

4.1 Continuity

The function f maps R^n into R^m if, for each vector or n-tuple $x \in R^n$, f associates with x a point $y = f(x) \in R^m$. In this case, we write $f: R^n \to R^m$.

EXAMPLE 4.1: Let $f(x) = \|x\|$ for $x \in R^n$. Then $f: R^n \to R^1$ is the function giving the length of the vector x.

EXAMPLE 4.2: $f_i(x_1, \ldots, x_n) = x_i$ is the i^{th} component function, f_i picks up the i^{th} component of $(x_1, \ldots, x_n)$.

EXAMPLE 4.3: $f: R^2 \to R^2$ satisfies $f(x_1, x_2) = (-x_2, x_1)$. This function rotates the axes 90^o counterclockwise, and all points with them.

DEFINITION 4.1: $f: R^n \to R^m$ is continuous at x_0 if $(\forall \varepsilon > 0)(\exists \delta > 0)$ $\|x - x_0\| < \delta \Rightarrow \|f(x) - f(x_0)\| < \varepsilon$.

Generally, we let $\| \;\|$ refer to the appropriate vector space. That is, if $x \in R^n$, then $\|x\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$. For any $f: R^n \to R^m$, the component functions $f_i: R^n \to R$ are defined by $f(x) = (f_1(x), f_2(x), \ldots, f_m(x))$. Thus, definition 4.1 defines continuity as requiring that as x gets close to x_0, f(x) gets close to $f(x_0)$. This mimics definition 2.6, with the only change that we must use norms instead of absolute value as our notion of distance. If f is continuous at all x_0, we say f is continuous.

THEOREM 4.1: f is continuous at x_0 if and only if all of the component functions of f are continuous at x_0.

Proof: $(\Rightarrow)$ Let $\varepsilon > 0$. Then $\exists \delta > 0$, $\|x - x_0\| < \delta \Rightarrow \|f(x) - f(x_0)\| < \varepsilon$.
Thus, if $\|x - x_0\| < \delta$

$|f_i(x) - f_i(x_0)| \le \|f(x) - f(x_0)\| < \varepsilon$

and f_i is continuous at x_0.

$(\Leftarrow)$ Suppose each f_i is continuous at x_0, and let $\varepsilon > 0$. Since $\varepsilon/\sqrt{n} > 0$, there is a $\delta_i > 0$

$\|x-x_0\| < \delta_i \Rightarrow |f_i(x) - f_i(x_0)| < \varepsilon.$

Thus, if $\|x-x_0\| < \delta = \min\{\delta_1, \delta_2,\ldots,\delta_n\}$

$$\|f(x) - f(x_0)\| = (\sum_{i=1}^{n}(f_i(x) - f_i(x_0))^2)^{½} \leq (\sum_{i=1}^{n}(\varepsilon/\sqrt{n})^2)^{½} = \varepsilon$$

and f is continuous. Q.E.D.

Thus $f:R^n \to R^m$ is continuous if and only if all of its components are continuous. Thus, functions into R^m are, as far as continuity is concerned, really only a grouping of m functions into R, or real valued functions.

THEOREM 4.2: $f:R^n \to R^m$ is continuous at x_0 if and only if, for all sequences $x^i \to x_0$, the sequence $f(x^i) \to f(x_0)$.

Proof: exercise 4.5.

This chapter contains several theorems whose proofs are exercises. Typically, the corresponding theorem in chapter 2 will provide the proof if $|\ |$ is replaced with $\|\ \|$.

THEOREM 4.3: Suppose $f:R^n \to R^m$ and $g:R^n \to R^m$ are continuous. Then

i). $h(x) = f(x) + g(x)$ is continuous

ii). $h(x) = \alpha f(x)$ is continuous for scalar α

iii). $h(x) = f(x) \cdot g(x)$ is continuous

Proof: exercise 4.6.

THEOREM 4.4: Suppose $f:R^n \to R^m$ and $g:R^m \to R^k$ are both continuous. Then $h:R^n \to R^k$ defined by $h(x) = g(f(x))$ is continuous.

Proof: Let $\varepsilon > 0$. Since g is continuous, $(\forall y_0)(\exists \delta_1 > 0)$

$\|y-y_0\| < \delta_1 \Rightarrow \|g(y)-g(y_0)\| < \varepsilon$

Since f is continuous at x_0, $\exists\ \delta_2 > 0$

$\|x-x_0\| < \delta_2 \Rightarrow \|f(x) - f(x_0)\| < \delta_1$

Thus, if $\|x-x_0\| < \delta_2$, letting $y = f(x)$ and $y_0 = f(y_0)$

$\|x-x_0\| < \delta_2 \Rightarrow \|f(x) - f(x_0)\| < \delta_1 \Rightarrow \|g(f(x)) - g(f(x_0))\| < \varepsilon.$

Thus $h(x) = g(f(x))$ is continuous at x_0. Since x_0 was arbitrarily chosen, h is continuous.

Q.E.D.

DEFINITION 4.2: $A \subseteq R^n$ is open if

$(\forall x \in A)(\exists \varepsilon > 0)\ \{y / \|x-y\| < \varepsilon\} \subseteq A$

If A^c is open, A is said to be closed.

It is useful to abbreviate

$N_\varepsilon(x) = \{y / \|x-y\| < \varepsilon\}$,

which is called an ε-ball or ε neighborhood. In R^2, $N_\varepsilon(x)$ is the set of points within a disk of radius ε, centered at x. In R^1, $N_\varepsilon(x)$ is the open interval $(x-\varepsilon, x+\varepsilon)$. In R^3, $N_\varepsilon(x)$ describes the set of points inside the sphere of radius ε centered at x.

In an open set A, all points are in the interior of A, in the sense that, at any point $x \in A$, we can move by an amount $\varepsilon > 0$ and still stay inside A. Thus, open sets do not contain their borders, for at the border of A, any movement away from A puts one outside A.

THEOREM 4.5: Let $A_i \subseteq R^n$ be open for all $i \in \Gamma$. Then

$\bigcup_{i \in \Gamma} A_i$ is open.

Also, if Γ has finitely many elements, $\bigcap_{i \in \Gamma} A_i$ is open. Finally, ϕ and R^n are open.

Proof: Suppose $x \in \bigcup_{i \in \Gamma} A_i$. Then, by Definition 1.2, $\exists j \in \Gamma$, $x \in A_j$. Since A is open, $\exists \varepsilon > 0$, $N_\varepsilon(x) \subseteq A_j \subseteq \bigcup_{i \in \Gamma} A_i$. Thus $N_\varepsilon(x) \subseteq \bigcup_{i \in \Gamma} A_i$. Since x was chosen arbitrarily, $\bigcup_{i\in\Gamma} A_i$ is open.

Now suppose Γ has finitely many elements. If $x\in \bigcap_{i\in\Gamma} A_i$, then $\forall i\in\Gamma$, $x\in A_i$. Since A_i is open, $\exists\ \varepsilon_i > 0$, $N_{\varepsilon_i}(x) \subseteq A_i$. Let $\varepsilon=\min\{\varepsilon_i/i\in\Gamma\}$. Since Γ is finite, $\varepsilon > 0$. Furthermore $N_\varepsilon(x) \subseteq N_{\varepsilon_i}(x) \subseteq A_i$ for all i, and thus, $N_\varepsilon(x) \subseteq \bigcap_{i\in\Gamma} A_i$. Since x was arbitrary, $\bigcap_{i\in\Gamma} A_i$ is open.

Finally, ϕ is open trivially, since there are no $x \in \phi$. R^n is open, since $N_1(x) \subseteq R^n$ for all x.

Q.E.D.

Theorem 4.5 shows that arbitrary unions and finite intersections of open sets are open. We see from the proof, as well, that typically intersections of infinitely many open sets are not open, and the reason is the $\varepsilon = \min\{\varepsilon_i/i\in\Gamma\}$ used in the proof may be zero.

EXAMPLE 4.4: Let $A_n = (-1/n, 1/n) \subseteq R$, and $\Gamma = \{1,2,3,\ldots\}$. Then

$$\bigcap_{i\in\Gamma} A_i = \{0\}$$

which is not open.

THEOREM 4.6: Let $A_i \subseteq R^n$ be closed for all $i \in \Gamma$. Then

$$\bigcap_{i\in\Gamma} A_i \text{ is closed.}$$

Also, if Γ has finitely many elements, $\bigcup_{i\in\Gamma} A_i$ is closed. ϕ and R^n are closed.

Proof: This is just Theorem 4.5, and De Morgan's Laws.

In words, arbitrary intersections, and finite unions, of closed sets are closed. The use of closed sets is that they contina their limit points.

THEOREM 4.7: Suppose $x_n \to x_0$, and $x_n \in A$ for all n, and A is closed. Then $x_0 \in A$.

Proof: By contradiction: suppose $x_0 \notin A$. Then $x_0 \in A^c$, which is open by definition. Thus $\exists \varepsilon > 0$ so that $N_\varepsilon(x_0) \subseteq A^c$. Since $x_n \to x_0$ $\exists N$, $n \geq N \Rightarrow \|x_n - x_0\| < \varepsilon$, which means $\forall n \geq N$, $x_n \in N_\varepsilon(x_0) \subseteq A^c$, so $x_n \notin A$, contradicting the hypothesis $x_n \in A$.

Q.E.D.

Thus, if A is closed, all cauchy sequences in A converge to something in A. Indeed, this characterizes the notion of a closed set, by the following theorem.

THEOREM 4.8: $A \subseteq R^n$ is closed if and only if all cauchy sequences x_n, $(\forall n)$ $x_n \in A$, converge to a point $x_o \in A$.

Proof: ($\Rightarrow$) From Theorem 3.17, all cauchy sequences converge, and by Theorem 4.7, the limit is in A.

($\Leftarrow$) By contrapositive, we show that if A is not closed, then there exists a cauchy sequence converging to something outside A.

Since A is not closed, A^c is not open, and $\exists\, x_0 \in A^c$ $\forall \varepsilon > 0$, $N_\varepsilon(x_0)$ is not contained in A^c. Thus, $\forall \varepsilon > 0$, $N_\varepsilon(x_0) \cap A \neq \phi$, (for $N_\varepsilon(x_0) \cap A = \phi \Rightarrow N_\varepsilon(x_0) \subseteq A^c$). Let $x_1 \in N_1(x_0) \cap A$ (since this is nonempty), and generally let

$$x_n \in N_{1/n}(x_0) \cap A$$

(x_n exists since $N_\varepsilon(x_0) \cap A$ is nonempty for all ε, and in particular $\varepsilon = 1/n$).

Clearly $x_n \to x_0$, and hence is cauchy. By construction, $x_n \in A$ for all n. Finally, $x_0 \notin A$ by construction.

Q.E.D.

Thus, we see that closed sets contain their limit points, i.e., if we construct a convergent sequence out of members of a closed set, it converges to something in the set.

Let $A \subseteq R^n$ and $f:R^n \to R^m$. We shall generally use the notation:

$$f(A) = \{y \in R^m / (\exists x \in A)\ f(x) = y\}$$

Thus, f(A) is the subset of R^m that f maps A into. Similarly, for $B \subseteq R^m$,

$$f^{-1}(B) = \{x \in R^n / f(x) \in B\}$$

$f^{-1}(B)$ is the subset of R^n that is mapped into B.

THEOREM 4.9: $f:R^n \to R^m$ is continuous if and only if, for all open sets $B \subseteq R^m$, $f^{-1}(B)$ is open.

Proof: ($\Rightarrow$) Suppose f is continuous and B is open. For each $f(x) \in B$, $\exists\ \varepsilon_x > 0$ and $N_{\varepsilon_x}(f(x)) \subseteq B$. Since f is continuous, at each point x,

$\exists\ \delta_x > 0$ so that $\|y-x\| < \delta_x \Rightarrow \|f(y)-f(x)\| < \varepsilon_x$.

or

$y \in N_{\delta_x}(x) \Rightarrow f(y) \in N_{\varepsilon_x}(f(x)) \subseteq B$.

Thus, if $y \in N_{\delta_x}(x)$, then $f(y) \in B$, forcing $\forall y \in N_{\delta_x}(x)$, $y \in f^{-1}(B)$, or

$$N_{\delta_x}(x) \subseteq f^{-1}(B).$$

Thus $f^{-1}(B)$ is open, since there is a neighborhood $N_{\delta_x}(x)$, around each point $x \in f^{-1}(B)$, that is contained in $f^{-1}(B)$.

($\Leftarrow$) Let $x \in R^n$ and $\varepsilon > 0$. Let

$$B = \{z \in R^m / \|z-f(x)\| < \varepsilon\}.$$

By hypothesis $f^{-1}(B)$ is open, and by definition, $x \in f^{-1}(B)$. Thus there is a $\delta > 0$, so that $N_\delta(x) \subseteq f^{-1}(B)$. That is,

$\forall y \in N_\delta(x)$, $f(y) \in B$, or, equivalently,

$\|x-y\| < \delta \Rightarrow \|f(x) - f(y)\| < \varepsilon$.

Thus f is continuous.

Q.E.D.

Theorem 4.9 shows the equivalence of the ε-δ definition of continuity and the effect of f on open sets (in particular, f^{-1} of an open set is open). The latter is called the topological definition of continuity, and, generally, many of our results can be expressed in terms of open sets. The field called topology proves theorems about open and closed sets in abstract spaces, and provides some elegant proofs of quite deep theorems. Topology has been more successfully applied in Economics (mostly in social choice) than in any other science, and we shall return to it in a later chapter.

4.2 MATRICES

A matrix is an array of real numbers. Matrices have dimensions mXn, meaning the array has m rows and n columns for integers m and n. Symbolically, an mXn matrix A may be represented

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{4.1}$$

EXAMPLE 4.5: For m=1, a matrix is a real vector in R^n, $(a_{11}, a_{12},\ldots,a_{1n})$, and is called a row vector since it is a row of a matrix. For n=1, the mX1 matrix is a column vector, being comprised of a single column

$$A = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}$$

EXAMPLE 4.6: Some matricies, with their dimensions listed below then

$$\underset{2X3}{\begin{pmatrix} 1 & \sqrt{2} & 16 \\ 4 & 3 & 1.9 \end{pmatrix}} \qquad \underset{2X2}{\begin{pmatrix} 4 & 8 \\ 7 & 3.1 \end{pmatrix}} \qquad \underset{2X1}{\begin{pmatrix} 1 \\ 7 \end{pmatrix}}$$

It is useful to refer to a typical component of a matrix, a_{ij} is the element in the i^{th} row and j^{th} column of A in 4.1. Matrices of the same dimensions mXn are added by adding the components:

$$A+B = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & & \vdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{pmatrix} \tag{4.2}$$

Rather than write out the entire matrix, we may refer to the i,j^{th} term

$$(A + B)_{i,j} = A_{i,j} + B_{i,j} = a_{ij} + b_{ij} \qquad (4.3)$$

Scalar multiplication of matrices is accomplished by multiplying all the components by the scalar

$$\lambda A = \begin{pmatrix} \lambda a_{11} & \dots & \lambda a_{1n} \\ \vdots & & \\ \lambda a_{m1} & \dots & \lambda a_{mn} \end{pmatrix} \qquad (4.4)$$

The reader may verify that the set of mXn matrices forms a vector space, given these definitions. Matrices with different dimensions cannot be added.

EXAMPLE 4.7

$$\begin{pmatrix} 4 & 7 \\ 3 & 1 \\ 5 & 5 \end{pmatrix} + \begin{pmatrix} 6 & 0 \\ 3 & 3 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 4+6 & 7+0 \\ 3+3 & 1+3 \\ 5+1 & 5+5 \end{pmatrix} = \begin{pmatrix} 10 & 7 \\ 6 & 4 \\ 6 & 10 \end{pmatrix}$$

$$2\begin{pmatrix} 4 & 7 \\ 3 & 1 \\ 5 & 5 \end{pmatrix} = \begin{pmatrix} 8 & 14 \\ 6 & 2 \\ 10 & 10 \end{pmatrix}$$

The transpose of a matrix A, denoted A^T, turns columns into rows and rows into columns.

$$A^T = \begin{pmatrix} a_{11} & a_{12} & \dots a_{1n} \\ a_{21} & a_{22} & \dots a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots a_{mn} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{21} & \dots a_{m1} \\ a_{12} & a_{22} & \dots a_{m2} \\ \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \dots a_{mn} \end{pmatrix} \qquad (4.5)$$

Thus, the transpose of an mXn matrix is nXm.

EXAMPLE 4.8

$$\begin{pmatrix} 3 & 7 \\ 2 & 2 \\ 8 & 9 \end{pmatrix}^T = \begin{pmatrix} 3 & 2 & 8 \\ 7 & 2 & 9 \end{pmatrix} \qquad \begin{pmatrix} 4 & 8 & 6 \\ 5 & 0 & 1 \end{pmatrix}^T = \begin{pmatrix} 4 & 5 \\ 8 & 0 \\ 6 & 1 \end{pmatrix}$$

In the typical element notation

$$(a_{ij})^T = a_{ji} \tag{4.6}$$

Matrices may be "multiplied" through a procedure which looks odd at first glance but is very useful. If A is mXn and B is nXk, the result of multiplying A and B, denoted AB, is mXk. The i,j^{th} element of AB is

$$(AB)_{ij} = \sum_{\ell=1}^{n} a_{i\ell} b_{\ell j} \tag{4.7}$$

This corresponds to taking the i^{th} row of A and the j^{th} column of B and applying the dot product. If we denote the i^{th} column of A by $a_{\cdot i}$ and the j^{th} row of A by $a_{j\cdot}$, so

$$a_{\cdot i} = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \cdot \\ \cdot \\ \cdot \\ a_{mi} \end{pmatrix} \qquad a_{j\cdot} = (a_{j1}, a_{j2}, \ldots, a_{jn}) \tag{4.8}$$

Then

$$(AB)_{ij} = a_{\cdot i} \cdot b_{j\cdot} \tag{4.9}$$

EXAMPLE 4.9

$$\begin{pmatrix} 3 & 2 & 0 \\ 1 & 1 & 5 \end{pmatrix} \begin{pmatrix} 6 & 1 \\ 3 & 3 \\ 2 & 5 \end{pmatrix} = \begin{pmatrix} 3 \cdot 6 + 2 \cdot 3 + 0 \cdot 2 & 3 \cdot 1 + 2 \cdot 3 + 0 \cdot 5 \\ 1 \cdot 6 + 1 \cdot 3 + 5 \cdot 2 & 1 \cdot 1 + 1 \cdot 3 + 5 \cdot 5 \end{pmatrix}$$

$$= \begin{pmatrix} 24 & 9 \\ 19 & 29 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 6 & 6 \\ 4 & 1 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \cdot 1 + 6 \cdot 5 + 6 \cdot 1 \\ 4 \cdot 1 + 1 \cdot 5 + 5 \cdot 1 \end{pmatrix} = \begin{pmatrix} 39 \\ 14 \end{pmatrix}$$

An nXn matrix is symmetric if $A^T = A$, that is, $a_{ij} = a_{ji}$. The reader should verify the following properties, for A, B mXn and C is nXk:

$$A + B = B + A \tag{4.11}$$

$$(A + B)C = AC + BC \tag{4.12}$$

$$(AC)^T = C^T A^T \tag{4.13}$$

$$(A^T)^T = A \tag{4.14}$$

Generally, AB is not equal to BA.

EXAMPLE 4.10

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & 2 \end{pmatrix} \neq \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

However, if A and B are both nXn and symmetric:

$$AB \overset{(4.10)}{=} A^T B^T \overset{(4.13)}{=} (BA)^T \tag{4.15}$$

since BA is symmetric if A and B are.

An important special case emerges when x is an nX1 column vector and A is an mXn matrix:

$$Ax = \begin{pmatrix} a_{\cdot 1} \cdot x \\ a_{\cdot 2} \cdot x \\ \vdots \\ a_{\cdot m} \cdot x \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} a_{i1} x_i \\ \vdots \\ \sum_{i=1}^{n} a_{im} \cdot x_i \end{pmatrix} \tag{4.16}$$

Thus, $f(x) = Ax$ is a function mapping R^n into R^m.

EXAMPLE 4.11: The function $f(x) = Ax$ where $x \in R^2$ and $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ rotates the axes by 90^o. To see this, note

$x \cdot Ax = (x_1,x_2) \cdot A(x_1,x_2) = (x_1,x_2) \cdot (-x_2,x_1) = -x_1x_2 + x_2x_1 = 0,$

and thus Ax is orthogonal (perpendicular) to x for any x. Thus A rotates all vectors by 90^o.

DEFINTION 4.3: An nXn matrix A is positive semidefinite (positive definite) if

$$(\forall x \neq \bar{0})\ x^T Ax \geq 0\ (> 0) \tag{4.17}$$

and is negative semidefinite (negative definite) if

$$(\forall x \neq \bar{0})\ x^T Ax \leq 0\ (< 0) \tag{4.18}$$

We see immediately that, if A is positive semidefinite, then the function $f(x) = Ax$ rotates every vector x by no more than 90^o, since $x^T Ax = x \cdot (Ax) \geq 0$, indicating the angle between x and Ax does not exceed 90^o. Analogously, positive definite matrices map vectors x into new vectors Ax less than 90^o away. Negative definite matrices map vectors more than 90^o away from their starting point.

DEFINITION 4.4: λ is an eigenvalue for an nXn matrix A with associated eigenvector $v \neq \bar{0}$ if

$$Av = \lambda v \tag{4.19}$$

Generally eigenvalues and eigenvectors can involve complex (imaginary, involving $\sqrt{-1}$) terms.

EXAMPLE 4.12: The matrix

$\begin{pmatrix} 2 & -3 \\ -1 & 4 \end{pmatrix}$ has eigenvalues 1 and 5 with eigenvectors $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Since

$$\begin{pmatrix} 2 & -3 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 & -3 \\ -3 + 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & -3 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 + 3 \\ -1 - 4 \end{pmatrix} = 5 \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

EXAMPLE 4.13: The matrix

$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ has eigenvalues $\sqrt{-1}$ and $-\sqrt{-1}$ and associated eigenvectors

$\begin{pmatrix} 1 \\ \sqrt{-1} \end{pmatrix}$ and $\begin{pmatrix} -1 \\ \sqrt{-1} \end{pmatrix}$.

From (4.19), we see that if v is an eigenvector, then αv is an eigenvector for any scalar $\alpha \neq 0$. An important matrix is the identity matrix

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \tag{4.20}$$

NOTE, for nXn matrix A

$$IA = AI = A \tag{4.21}$$

If A is an nXn matrix, A^{-1} is defined by

$$AA^{-1} = I \tag{4.22}$$

if such a matrix A^{-1} exists.

EXAMPLE 4.14:

if $A = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$, then $A^{-1} = \begin{pmatrix} 1/3 & -1/3 \\ +1/3 & 2/3 \end{pmatrix}$

If $A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}$, then A^{-1} doesn't exist.

THEOREM 4.10: The following are equivalent, for nXn symmetric matrix A:

i. A^{-1} exists

ii. no eigenvalue of A is zero

iii. $\{y \in R^n / (\exists x \in R^n)\ Ax = y\} = R^n$

iv. All the column vectors of A are linearly independent.

v. All the row vectors of A are linearly independent.

Theorem 4.10 is stated without proof. However, an understanding of its implications arises from remembering that the nXn matrix A also defines a function $f:R^n \to R^n$ by $f(x) = Ax$. Thus, for A^{-1} to exist A must be 1-1 and onto. Part iii. says A is onto. If all of the columns of A are linearly independent, then the n vectors Ae_i, where

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad i^{th} \text{ component} \tag{4.23}$$

form a linearly independent set of dimension n, since each one is merely a column of A. Thus part iv. says A is onto. Part v. says the same about A^T, an equivalent problem since

$$(A^T)^{-1} = (A^{-1})^T$$

if A^{-1} exists.

Let A be an nXn symmetric matrix with eigenvalues $\lambda_1,\ldots,\lambda_n$. They will all be real in this case, since A is symmetric. If A is positive definite, then $x^T Ax > 0$ for $x \neq 0$. This is, in particular, true for any eigenvector v_i. Thus $0 < v_i^T A v_i = v_i^T \lambda_i v_i = \lambda_i(v_i \cdot v_i)$, and this shows $\lambda_i > 0$, since $v_i \cdot v_i > 0$. Thus, positive definite matrices have positive eigenvalues, positive semidefinite matrices have nonnegative eigenvalues, negative definite matrices have negative eigenvalues, and, finally, negative semidefinite matrices have nonpositive eigenvalues. These arguments go the other direction as well (in the case of a square, symmetric, invertible matrix), because generally we can write any vector x as a linear combination of the eigenvectors

$$x = \sum_{i=1}^{n} \alpha_i v_i$$

so that $$x^T Ax = \sum_{i=1}^{n} \alpha_i^2 \lambda_i (v_i \cdot v_i)$$

which is positive for all x if and only if all the eigenvalues λ_i are positive.

Generally, square matrices perform a job of rotating and expanding or contracting vectors. Since we can write

$$Ax = \sum_{i=1}^{n} \alpha_i \lambda_i v_i$$

Ax moves vectors (except the eigenvectors) around, and it either lengthens ($|\lambda| > 1$) or shortens ($|\lambda| < 1$) then. Thus, if we use the eigenvectors as a basis for R^n, A serves the role of only increaseng or decreasing the length of the vectors.

When a rotation of the space is performed, as in Example 4.11, there are no vectors which Ax maps to λx (since they are all rotated away), at least in R^n, and in this case, complex eigenvalues arise. Generally, with a complex eigenvalue $a + b\sqrt{-1}$, the real part a describes the expansion or contraction of the vector, while the imaginary part b describes the rotation of the whole space. This will be further discussed in Chapter ?.

DEFINITION 4.5: $f: R^n \to R^m$ is a linear function if $\forall x, y \in R^n$ and scalars α, β

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \tag{4.24}$$

THEOREM 4.11: $f: R^n \to R^m$ is a linear function if and only if there is a mXn matrix A, $f(x) = Ax$.

Proof: ($\Rightarrow$) Let $a_{\cdot i} = f(e_i)$, that is, the i^{th} column of A is $f(e_i)$. Then $f(x) = f(x_1, \ldots, x_n) = f(x_1 e_1 + \ldots + x_n e_n) = x_1 f(e_1) + \ldots + x_n f(e_n) =$

$$\begin{pmatrix} a_{11} x_1 \\ a_{21} x_1 \\ \vdots \\ a_{m1} x_1 \end{pmatrix} + \ldots + \begin{pmatrix} a_{1n} x_n \\ a_{2n} x_n \\ \vdots \\ a_{mn} x_n \end{pmatrix} = \begin{pmatrix} a_{11} x_1 + \ldots + a_{1n} x_n \\ a_{21} x_1 + \ldots + a_{2n} x_n \\ \vdots \\ a_{m1} x_1 + \ldots + a_{mn} x_n \end{pmatrix} =$$

Ax.

($\Leftarrow$) Exercise.

Thus, the linear functions are precisely those that may be represented by matrices.

The theory of derivatives of real functions makes great use of the notion of a tangency, that is, a tangent line to a function f at x is the straight line passing through (x, f(x)) with the same slope as f at this point. To generalize this notion, we must consider tangent planes. Consider a function $f: R^2 \to R$, and visualize $f(x_1 x_2)$ as altitude; $f(x_1, x_2)$ is the height of a smooth piece of land (rolling hills, for example). At any point on the land's surface, we can take a plane (a piece of plywood) and make it tangent at this point.

The general description of a plane in R^3 is

$a_1 x_1 + a_2 x_2 + a_3 x_3 = b$

for scalars a_1, a_2, a_3 and b. If $a_3 \neq 0$, we can write this as

$$x_3 = \frac{b}{a_3} - \frac{a_1}{a_3} x_1 - \frac{a_2}{a_3} x_2 .$$

Thus b/a_3 is the planes "height" over the (x_1, x_2) axes (the plant $x_3 = 0$), $-a_1/a_3$ is the slope in the x_1 direction and $-a_2/a_3$ is the slope in the x_2 direction.

Thus, if $\alpha \in R^3$ and $b \in R$, we may express the formula for a plane as

$\alpha \cdot x = b.$

Where $\alpha = (a_1, a_2, a_3)$ above. In an analogous way, the formula for a line in R^2 is

$\alpha_1 x_1 + \alpha_2 x_2 = b$, or

$\alpha \cdot x = b.$

Thus, generally a "plane" in a higher dimensional space is described by the formula

$$\alpha \cdot x = b$$

for constants $\alpha \in R^n$ and $b \in R$. The parameter b is the height of the plane over $(0,\ldots,0) = \bar{0}$, while α_i is the slope of the plane in the direction x_i. Such "planes" are called hyperplanes, to distinguish them from ordinary planes in R^3.

Consider any two points x and y on a hyperplane. Then

$$\alpha \cdot x = b$$

$$\alpha \cdot y = b.$$

Subtracting:

$$\alpha \cdot (x-y) = 0.$$

Thus, the vector α is perpendicular to any line segment x-y on the plane. Thus, a plane is determined by a vector (which every line in the plane is perpendicular to) and its height over a point (figure 1).

Suppose $f: R^n \to R$ and we wish to approximate f, at the point x^0, by a tangent hyperplane, that is, by a hyperplane in R^{n+1} whose "height" (last coordinate) over x^o equals $f(x^0)$, and whose slopes in each direction are the same at x^0. Thus, the hyperplane must satisfy

$$x_{n+1} = \alpha \cdot x + b \qquad \text{(definition of hyperplane)}$$

$$= \alpha \cdot (x-x^0) + \beta \qquad \beta = b + (\alpha \cdot x^0)$$

$$= \alpha \cdot (x-x^0) + f(x_0) \qquad \text{(for height to be the same at } x_0\text{).}$$

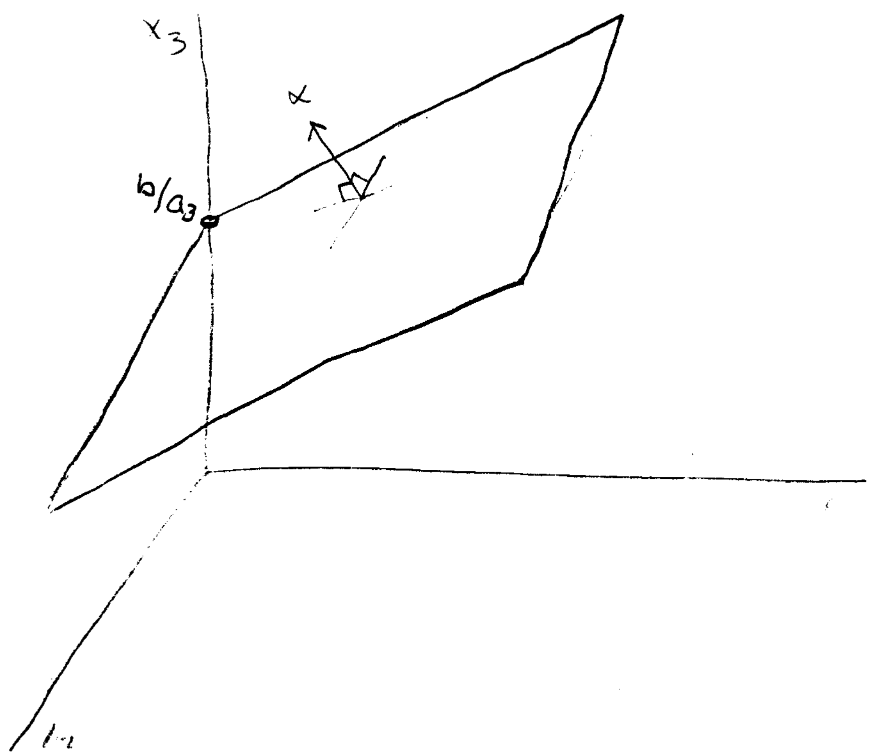

Figure 4.1 : The plane given by $\alpha \cdot X = 0$.
α is perpendicular to the plane.

For the slopes to match up, α_i must be the slope of f at x as we change x_i only:

$$\alpha_i = \lim_{\lambda \to 0} \frac{f(x_1^0, \ldots x_{i-1}^0, x_i^0 + \lambda, x_{i+1}^0, \ldots x_n^0) - f(x^0)}{\lambda}$$

$$= \frac{\partial f}{\partial x_i}(x^0)$$

Thus, as we shall develop more carefully in the next section, the derivative of a function $f:R^n \to R$ defines the tangent hyperplane to f at x_0, and is essentially a linear approximation to f at x_0. The hyperplane, coming in the form $\alpha \cdot x = b$, is such that α is perpendicular to the hyperplane. This provides a geometric intuition to the results of the next section.

4.3 DERIVATIVES

If $f:R \to R$, we have identified $f'(x^0)$ with the slope of f at x^0. Thus, if $f:R^n \to R$, we can consider the slope of f as we vary one of the components of f. For example, let $g:R \to R$ be defined by

$$g(x_i) = f(x_1^0, \ldots, x_{i-1}^0, x_i, x_{i+1}^0, \ldots, x_n^0) \tag{4.25}$$

Then $g:R \to R$ is merely the function f, holding $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ constant at the values of x^0 while letting the i^{th} component vary. We can take the derivative of g in the normal way, and this leads to the notion of a partial derivative, partial because we are accomplishing only part of the differentiating of f (with respect to one component).

DEFINITION 4.6: The partial derivative of $f:R^n \to R$ with respect to x_i, at $x \in R^n$, is

$$D_i f(x) = \frac{\partial f(x)}{\partial x_i} = \lim_{\substack{\lambda \to 0 \\ \lambda \neq 0}} \frac{f(x + \lambda e_i) - f(x)}{\lambda} \tag{4.26}$$

if this limit exists.

If all the partial derivatives exist, the gradient of f is the vector of these partials:

$$\nabla f(x) = (\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)) \tag{4.27}$$

The gradient of f is called "del f" on occasion. Since $\frac{\partial f}{\partial x_i}$ is the slope of f in the direction e_i, the gradient summarizes the information about how f is changing in the n directions $e_1, \ldots, e_n$. In a similar fashion, we may take a derivative in any direction $y \in R^n$ provided $y \neq \bar{0}$. That is, we consider the function $g:R \to R$ given by

$$g(\lambda) = f(x_0 + \lambda y) \tag{4.28}$$

so that $g(\lambda)$ is the value of f as we move in the direction y from x_0. g is an ordinary function of a real variable, so its derivative can be defined in the ordinary way. Clearly $g'(0)$ is the slope of f as we move in the direction y from x_0.

DEFINITION 4.7: The directional derivative of $f:R^n \to R$ the direction $y \neq \bar{0}$ is

$$f_y(x_0) = \lim_{\substack{\lambda \to 0 \\ \lambda \neq 0}} \frac{f(x_0 + \lambda y) - f(x_0)}{\lambda} \tag{4.29}$$

if this limit exists.

EXAMPLE 4.15: Define $f:R^n \to R$ by

$$f(x_1,x_2) = \begin{cases} 1 & x_1 > 0 \text{ and } x_2 = x_1^2 \\ 0 & \text{otherwise} \end{cases}$$

See figure 4.2. f has directional derivatives in all directions at (0,0), but is discontinuous at (0,0).

This example shows that, even if all of the directional derivatives are defined for every direction $y \neq \bar{0}$, a function may fail to be continuous. Thus, in some sense, the existence of directional derivatives is not very useful, since it fails to guarantee continuity. A more productive way to think about derivatives results from thinking of derivatives as local linear approximations to f. If $f:R \to R$, then for some δ

$$|x-x_0| < \delta \Rightarrow \frac{|f(x) - f(x_0) - f'(x_0)(x - x_0)|}{|x - x_0|} < \varepsilon \qquad (4.30)$$

That is, $f'(x_0)(x-x_0)$ approximates $f(x) - f(x_0)$ if x is close enough to x_0. $f'(x_0)(x-x_0)$ is a linear function of $(x-x_0)$, and thus $f'(x_0)$ represents the approximation of f by a linear function. If, instead, $x \in R^n$, a linear function mapping R^n into R is represented by a n X 1 matrix (or vector) from theorem 4.11. Thus, we can define the derivative of $f:R^n \to R$ as a linear function $L(x-x_0) = f'(x_0) \cdot (x-x_0)$

$$\lim_{\substack{x \to x_0 \\ \|x-x_0\|>0}} \frac{|f(x) - f(x_0) - f'(x_0) \cdot (x-x_0)|}{\|x-x_0\|} = 0 \qquad (4.31)$$

if this limit exists. If $f'(x_0)$ exists for all x_0, f is said to be differentiable.

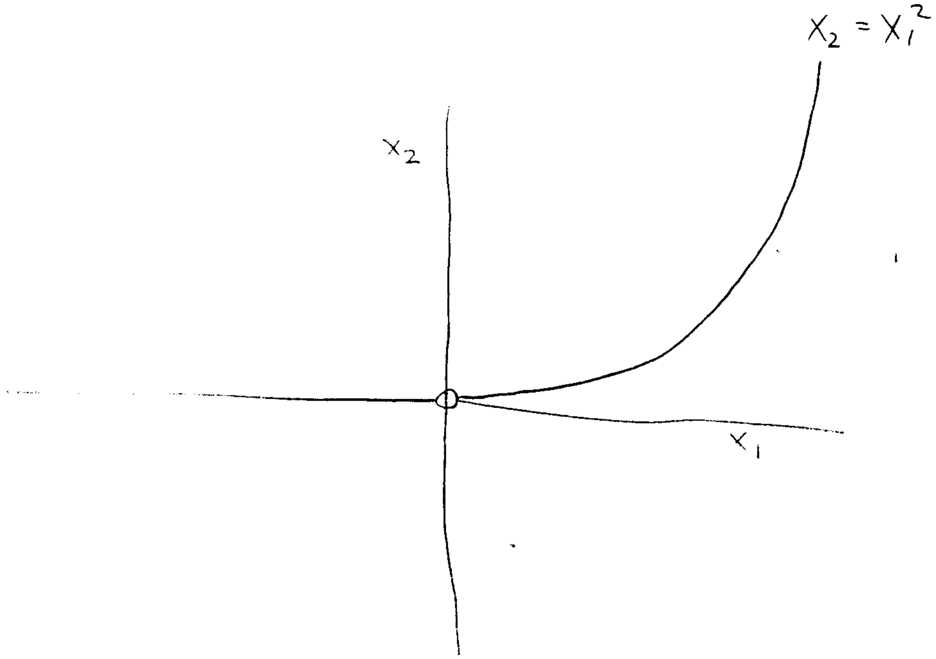

Figure 4.2 : f is 1 on $x_2 = x_1^2$, and zero everywhere else.

The notions of partial, directional and total derivatives ($f'(x_0)$ is said to be the total derivative) are the same if $f:R \to R$, since there is only one direction in R. Note that, if $f:R^n \to R$ is differentiable, the derivative of f is itself a function $f'(x_0)$ and $f':R^n \to R^n$, since for each $x_0 \in R^n$, $f'(x_0) \in R^n$.

EXAMPLE 4.16: $f(x_1,x_2) = x_1^2 + 3x_1x_2 + x_2^2$

$$\frac{\partial f}{\partial x_1}(x_1,x_2) = 2x_1 + 3x_2$$

$$\frac{\partial f}{\partial x_2}(x_1,x_2) = 3x_1 + 2x_2$$

If $y = (y_1,y_2)$ is any direction.

$$f(x + \lambda y) = (x_1 + \lambda y_1)^2 + 3(x_1 + \lambda y_1)(x_2 + \lambda y_2) + (x_2 + \lambda y_2)^2$$

and

$$\begin{aligned} f_y(x) &= 2x_1y_1 + 3x_2y_1 + 3x_1y_2 + 2x_2y_2 \\ &= (2x_1 + 3x_2,\ 3x_1 + 2x_2)\cdot(y_1,y_2) \\ &= \nabla f(x)\cdot y \end{aligned}$$

$f'(x) = \nabla f(x)$.

This example illustrates the results of the following theorem:

THEOREM 4.12: Suppose $f:R^n \to R$ is differentiable. Then all directional derivatives f_y exist, and

$$f'(x) = \nabla f(x) \tag{4.32}$$

$$f_y(x) = f'(x)\cdot y \tag{4.33}$$

Proof: Since

$$\frac{\partial f}{\partial x_i}(x) = f_{e_i}(x) \tag{4.34}$$

it is sufficient to prove (4.33), and this implies (4.32). So fix a direction y, and let $x = x_0 + \lambda y$. Then, by (4.31)

$$\lim_{\substack{\lambda\to 0\\ \lambda\neq 0}} \frac{|f(x_0 + \lambda y) - f(x_0) - f'(x_0)\cdot(\lambda y)|}{\|\lambda y\|} = 0$$

or, multiplying by $\|y\|$:

$$\lim_{\substack{\lambda\to 0\\ \lambda\neq 0}} \frac{|f(x_0 + \lambda y) - f(x_0) - \lambda f'(x_0)\cdot y|}{|\lambda|} = 0$$

It follows that

$$f_y(x_0) = \lim_{\substack{\lambda\to 0\\ \lambda\neq 0}} \frac{f(x_0 + \lambda y) - f(x_0)}{\lambda} = f'(x_0)\cdot y$$

Proving (4.33). (4.32) follows since

$$\frac{\partial f}{\partial x_i}(x_0) = f_{e_i}(x_0) = f'(x_0)\cdot e_i$$

Q.E.D.

Theorem 4.7 relates all of the notions of derivatives introduced so far. If f is differentiable, $f'(x)$ is just the vector of partial derivatives $\nabla f(x)$. In addition, the derivative of f in the direction y is $f'(x)\cdot y$. Because this changes scale with y (i.e. doubling y doubles the directional derivative), it is sometimes useful to normalize for $\|y\| = 1$, i.e. define the directional derivative in the direction y as $f'(x)\cdot(y/\|y\|)$ so that changes in scale do not affect the directional derivative.

When n=1, we saw that if f is differentiable, then f is continuous. This remains true.

THEOREM 4.13: If $f:R^n \to R$ is differentiable at x_0, then f is continuous at x_0.

Proof: Let $\varepsilon > 0$ and $h = x-x_0$. From (4.13), $\exists\delta_1 > 0$

$$0 < \|h\| < \delta_1 \Rightarrow \frac{|f(x_0+h) - f(x_0) - f'(x_0)\cdot h|}{\|h\|} < 1 \qquad (4.35)$$

Let $\delta = \min\{\delta_1, \frac{\varepsilon}{1 + \|f'(x_0)\|}\}$

Then, if $\|x-x_0\| = \|h\| < \delta$

$|f(x) - f(x_0)| = |f(x_0 + h) - f(x_0) - f'(x_0)\cdot h + f'(x_0)\cdot h| \leq$

$|f(x_0 + h) - f(x_0) - f'(x_0)\cdot h| + |f'(x_0)\cdot h| \leq$

$\|h\| + \|f'(x_0)\|\|h\| =$ (by 4.35 and cauchy schwarz)

$(1 + \|f'(x_0)\|)\|h\| < \varepsilon$ (since $\|h\| < \delta$).

Q.E.D.

This analysis is insufficient to differentiate the derivative of $f:R^n \to R$, since $f':R^n \to R^n$. Thus, we should generally like a definition of derivative which allows us to differentiate any function $f:R^n \to R^m$. We use our analysis of linear functions to allow this extension.

DEFINITION 4.9: The derivative of $f:R^n \to R^m$ at a point x_0 is a linear function $L:R^n \to R^m$ satisfying

$$\lim_{\substack{x \to x_0 \\ x \neq x_0}} \frac{\|f(x) - f(x_0) - L(x-x_0)\|}{\|x - x_0\|} = 0 \qquad (4.37)$$

if this exists. If f is differentiable at all x_0, f is said to be differentiable.

Since linear functions have matrix representations, we can represent L in (4.37) as an mXn matrix $f'(x_0)$. Let $f_i: R^n \to R$ be the component functions of f, so that

$$f(x) = (f_1(x),\ f_2(x),\ldots,f_m(x)) \tag{4.38}$$

THEOREM 4.14: f is differentiable if and only if all the component functions of f are differentiable, and in this case

$$f'(x) = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_2(x)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \frac{\partial f_m(x)}{\partial x_2} & \cdots & \frac{\partial f_m(x)}{\partial x_n} \end{pmatrix} \tag{4.39}$$

$f'(x)$ is called the Jacobean of f at x.

Proof: Let the mXn matrix A represent $f'(x_0)$, and $a_{i\cdot}$ the rows of A. Then, by Theorem 3.16,

$$\frac{f(x) - f(x_0) - A(x - x_0)}{\|x - x_0\|} \text{ converges to } \bar{0}$$

if and only if, $(\forall_i)$

$$\frac{f_i(x) - f_i(x_0) - a_{i\cdot}\bullet(x - x_0)}{\|x - x_0\|} \text{ converges to } 0.$$

Thus $A = f'(x_0)$ exists if and only if

$$(\forall_i) \lim_{\substack{x \to x_0 \\ x \neq x_0}} \frac{|f_i(x) - f_i(x_0) - a_{i\bullet} \cdot (x - x_0)|}{\|x - x_0\|} = 0 \qquad (4.40)$$

That is, $f'(x_0)$ exists if and only if all the component functions f_i are differentiable. Comparing (4.40) and (4.31), we see $a_{i\bullet}$ is ∇f_i, by (4.32), or

$$a_{ij} = \frac{\partial f_i}{\partial x_j}(x_0).$$

Q.E.D.

THEOREM 4.15: if $f:R^n \to R^m$ is defined by $f(x) = Ax$ for an mXn matrix A, then $f'(x) = A$.

Proof: We show A satisfies definition 4.8.

$$\lim_{\substack{x \to x_0 \\ x \neq x_0}} \frac{\|f(x) - f(x_0) - A(x - x_0)\|}{\|x - x_0\|} = \lim_{\substack{x \to x_0 \\ x \neq x_0}} \frac{\|Ax - Ax_0 - A(x-x_0)\|}{\|x - x_0\|}$$

$$= \lim_{\substack{x \to x_0 \\ x \neq x_0}} \frac{\|\bar{0}\|}{\|x-x_0\|} = 0. \qquad \text{Q.E.D.}$$

EXAMPLE 4.17: Let $f:R^2 \to R^2$ be given by

$$f(x_1,x_2) = (x_1^2 - x_1x_2 + x_2^2,\ x_1^2 - x_2^2)$$

$$f'(x_1,x_2) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 & 2x_2 - x_1 \\ 2x_1 & 2x_2 \end{pmatrix}$$

THEOREM 4.16: If $f:R^n \to R^m$ is differentiable at x_0, then f is continuous at x_0.

Proof: Follows immediately from Theorem 4.14 and Theorem 4.1.

Q.E.D.

THEOREM 4.17: Suppose $f:R^n \to R^m$, $g:F^n \to R^m$ are both differentiable. Then $h:R^n \to R^m$ defined by $h(x) = f(x) + g(x)$ is differentiable, and $h'(x) = f'(x) + g'(x)$.

Proof: Exercise.

THEOREM 4.18: Suppose $f:R^n \to R^m$ is differentiable and $g:R^m \to R^k$ is differentiable. Then $h:R^n \to R^k$ defined by $h(x) = g(f(x))$ is differentiable, and $h'(x) = g'(f(x))f'(x)$.

Theorem 4.18 shows the value in the definition of matrix multiplication. To find the derivative of the composite function g(f(x)), are merely multiples the matrices g'(f(x)) and f'(x). This treatment of derivatives as local linear approximations

$$f(x) - f(x_0) \doteq f'(x_0)(x-x_0) \qquad (4.41)$$

where $\doteq$ means approximately equal to, allows a geometric intuition, when m=n, so that $f:R^n \to R^n$. In this case, $f(x) - f(x_0)$ is parallel to $x-x_0$ whenever $f(x) - f(x_0) = \lambda(x-x_0)$ for some λ, but this requires

$$f'(x_0)(x-x_0) \doteq f(x) - f(x_0) = \lambda(x-x_0)$$

and thus $x-x_0$ is an eigenvector of $f'(x_0)$. It follows that we may think of f as, at least locally for x near x_0, as rotating and perhaps magnifying or diminishing vectors $x-x_0$, in the sense that $f(x) - f(x_0)$ is roughly a linear function of $x-x_0$.

Recall that, if $f:R^n \to R$ is differentiable, then its derivative f'(x) at a point x is itself a vector in R^n, that is, $f':R^n \to R^n$. Thus, if f' has a derivative, which we'll denote by f"(x), this derivative is an nXn matrix. We say f is twice continuously differentiable if every element of f"(x) is a continuous function of x. Since $f'(x) = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})$, the i, jth

component of f"(x) (ith row, jth column) is

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}(x)\right) = \frac{\partial^2 f(x)}{\partial x_j \partial x_i} \tag{4.42}$$

However, if f' is continuous, then f" is symmetric (when it exists):

$$\frac{\partial^2 f(x)}{\partial x_j \partial x_i} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \tag{4.43}$$

or, equivalently:

$$f''(x)^T = f''(x). \tag{4.44}$$

f"(x) is called a Hessian matrix.

Since $\frac{\partial^2 f(x)}{\partial x_j \partial x_i}$ is defined by

$$\frac{\partial^2 f(x)}{\partial x_j \partial x_i} = \lim_{\substack{\alpha\to 0\\ \alpha\neq 0}} \frac{1}{\alpha}\left(\frac{\partial f}{\partial x_i}(x + \alpha e_j) - \frac{\partial f}{\partial x_i}(x)\right)$$

$$= \lim_{\substack{\alpha\to 0\\ \alpha\neq 0}} \frac{1}{\alpha}\left(\lim_{\substack{\beta\to 0\\ \beta\neq 0}} \frac{1}{\beta}(f(x + \alpha e_j + \beta e_i) - f(x + \alpha e_j)) - \right.$$

$$\left. \lim_{\substack{\beta\to 0\\ \beta\neq 0}} \frac{1}{\beta}(f(x + \beta e_i) - f(x))\right)$$

$$= \lim_{\substack{\alpha\to 0\\ \alpha\neq 0}} \lim_{\substack{\beta\to 0\\ \beta\neq 0}} \frac{1}{\alpha\beta}(f(x + \alpha e_j + \beta e_i) - f(x + \alpha e_j) -$$

$$f(x + \beta e_i) + f(x))$$

$$= \lim_{\substack{\beta\to 0\\ \beta\neq 0}} \lim_{\substack{\alpha\to 0\\ \alpha\neq 0}} \frac{1}{\alpha\beta} (f(x + \alpha e_j + \beta e_i) - f(x + \beta e_i)$$

$$- (f(x + \alpha e_j) - f(x))$$

$$= \lim_{\substack{\beta\to 0\\ \beta\neq 0}} \frac{1}{\beta} \lim_{\substack{\alpha\to 0\\ \alpha\neq 0}} [\frac{1}{\alpha}(f(x + \beta e_i + \alpha e_j) - f(x + \beta e_i))$$

$$- \frac{1}{\alpha} (f(x + \alpha e_j) - f(x))]$$

$$= \lim_{\substack{\beta\to 0\\ \beta\neq 0}} \frac{1}{\beta} [\frac{\partial f}{\partial x_j} (x + \beta e_i) - \frac{\partial f}{\partial x_j}(x)] = \frac{\partial^2 f}{\partial x_i \partial x_j} \qquad (4.45)$$

THEOREM 4.19: Suppose $f:R^n \to R$ is twice differentiable, and define $g:R\to R$ by

$$g(\lambda) = f(x + \lambda y) \qquad (4.46)$$

Then $g'(\lambda) = f'(x + \lambda y)\cdot y$ (4.47)

and $g''(\lambda) = y^T f''(x + \lambda y)y$ (4.48)

Proof: (4.47) follows immediately from (4.29) and Theorem 4.12.

$$g''(\lambda) = \lim_{\substack{\alpha\to 0\\ \alpha\neq 0}} \frac{1}{\alpha}(g'(\lambda + \alpha) - g'(\lambda))$$

$$\overset{(4.47)}{=} \lim_{\substack{\alpha\to 0\\ \alpha\neq 0}} \frac{1}{\alpha}(y\cdot f'(x + \lambda y + \alpha y) - y\cdot f'(x + \lambda y))$$

$$= y\cdot \lim_{\substack{\alpha\to 0\\ \alpha\neq 0}} \frac{1}{\alpha}[f'(x + \lambda y + \alpha y) - f'(x + \lambda y)]$$

$$= y\cdot(f''(x + \lambda y)y) = y^T f''(x + \lambda y)y$$

Q.E.D.

Thus, as an immediate consequence of our second order approximation of real valued functions

$$g(1) = g(0) + g'(0)1 + \tfrac{1}{2}g''(\beta)1^2$$

for some $0 \leq \beta \leq 1$, we have

$$f(x+y) = f(x) + f'(x)\cdot y + \tfrac{1}{2}y^T f''(x + \beta y)y$$

This required g" to be continuous. Thus, we have shown:

THEOREM 4.20: Suppose $f:R^n \to R$ is twice continuously differentiable. Then there is a $\beta \in [0,1]$ so that

$$f(z) = f(x) + f'(x)\cdot(z-x) + \tfrac{1}{2}(z-x)^T f''(\beta z + (1-\beta)x)(z-x)$$

(β depends generally on x and z).

DEFINITION 4.10: $f:R^n \to R$ is concave if $\forall x, y \in R^n$ and $\forall \lambda \in [0,1]$

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$$

This definition compares immediately with Definition 2.8, as capturing the same notion. As a result, we obtain the analogous Theorem:

THEOREM 4.21: Suppose $f:R^n \to R$ is twice continuously differentiable. Then the following are equivalent:

i). f is concave

ii). $f(x) \leq f(y) + f'(y)\cdot(x-y)$

iii). $f''(y)$ is negative semidefinite

Proof: This theorem's proof is virtually identical to the proof of Theorem 2.23

DEFINITION 4.11: $f:R^n \to R$ is convex if $\forall x, y \in R^n$ and $\lambda \in [0,1]$

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

THEOREM 4.22: If $R^n \to R$ is twice continuously differentiable, the following are equivalent:

i). f is convex

ii). $-f$ is concave

iii). $f(x) \geq f(y) + f'(y) \cdot (x-y)$

iv). f'' is positive semidefinite

Proof: (i) iff (ii) follows directly from the definitions 4.9 and 4.10, while the equivalence of (iii) and (iv) follows immediately from Theorem 4.16.

Q.E.D.

One interesting aspect of concave or convex functions is the sets they define.

DEFINITION 4.12: $A \subseteq R^n$ is convex is $\forall x, y \in A$ and $\forall \lambda \in [0,1]$, $\lambda x + (1-\lambda)y \in A$.

Thus, the set A is convex if, for all x and y in A, the line segment connecting x and y is also in A (see figure 4.3 a,b). $\lambda x + (1-\lambda)y$ is called a convex combination of x and y for $0 \leq \lambda \leq 1$.

EXAMPLE 4.18: Recall that, if $x \in R^n$ represents consumption and $p \in R^n$ represents prices, a person with income y to spend can purchase any bundle x satisfying $p \cdot x \leq y$. Show $\{x/p \cdot x \leq y\}$ is a convex set.

Convex sets are intimately related to concave (and convex) functions, as the next theorem shows.

THEOREM 4.23: If $f:R^n \rightarrow R$ is concave, then $(\forall b \in R)$ $\{x/f(x) \geq b\}$ is convex. If $f:R^n \rightarrow R$ is convex, $\{x/f(x) \leq a\}$ is convex for all $a \in R$.

Proof: Let $x, y \in \{x/f(x) \geq b\}$ so that $f(x) \geq b$ and $f(y) \geq b$. Then, for $0 \leq \lambda \leq 1$:

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y) \geq \lambda b + (1-\lambda)b = b$$

so $\lambda x + (1-\lambda)y \in \{x/f(x) \geq b\}$ as desired. The second assertion has a similar proof.

Q.E.D.

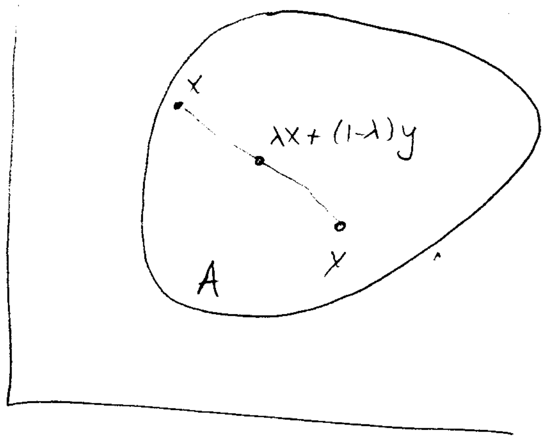

figure 4.3 a: $\lambda x + (1-\lambda) y \in A$, A convex

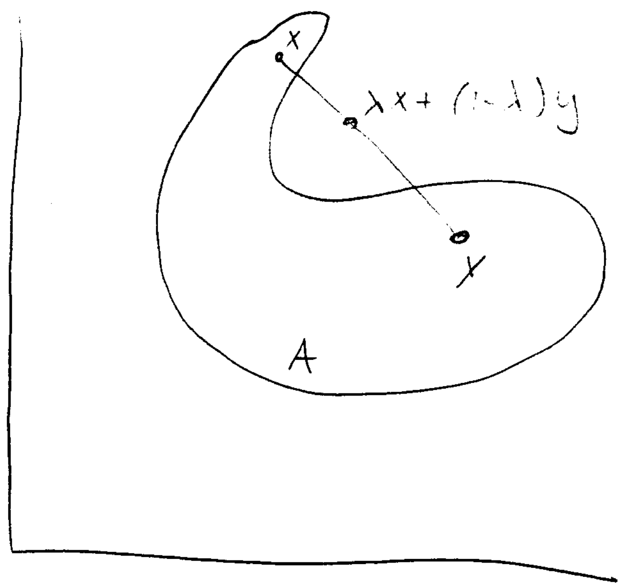

figure 4.3 b: $\lambda x + (1-\lambda) y \notin A$, A is not convex

Consider any $f:R^n \to R$. In what direction is $f(x)$ constant, starting from a point x_0? This requires, as λ gets small:

$$f(x_0 + \lambda z) = f(x_0)$$

or $$\lim_{\lambda \to 0} \frac{1}{\lambda}(f(x_0 + \lambda z) - f(x_0)) = 0$$

or $$f'(x_0) \cdot z = 0$$

That is, the surface $\{x/f(x) = f(x_0)\}$ is defined by orthogonality to $f'(x_0)$ (see figure 4.4). Or, put another way, the level surface $\{x/f(x) = a\}$ is perpendicular to $f'(x)$, that is, $f'(x)$ points directly away from any vector z so that $f(x + \lambda z)$ is approximately constant (as $\lambda \to 0$). Thus $f'(x_0) \cdot (x - x_0) + f(x_0)$ defines the tangent hyperplane for f at x_0.

This analysis extends to increases or decreases in f. When does a small movement in the direction z increase f? Whenever

$$\frac{d}{d\lambda} f(x + \lambda z) \bigg|_{\lambda=0} \geq 0$$

or

$$f'(x) \cdot z \geq 0.$$

Thus, small movements in directions that are within 90° of $f'(x)$ increase f, while movements in directions more than 90° from $f'(x)$ decrease f (figure 4.5). Thus $f'(x)$ points in the direction of increasing f. If $f:R^2 \to R$ describes a hill ($f(x_1, x_2)$ is the height of the hill, given coordinates (x_1, x_2)) then the vector f' points up the hill (toward the peak). The level curve $\{x/f(x) = a\}$ is the set of points of altitude a, and, as was argued, if $f'(x) \cdot z = 0$, then heading from x in the direction z means staying at a constant altitude. Generally, we see that directional derivatives $f'(x) \cdot y$

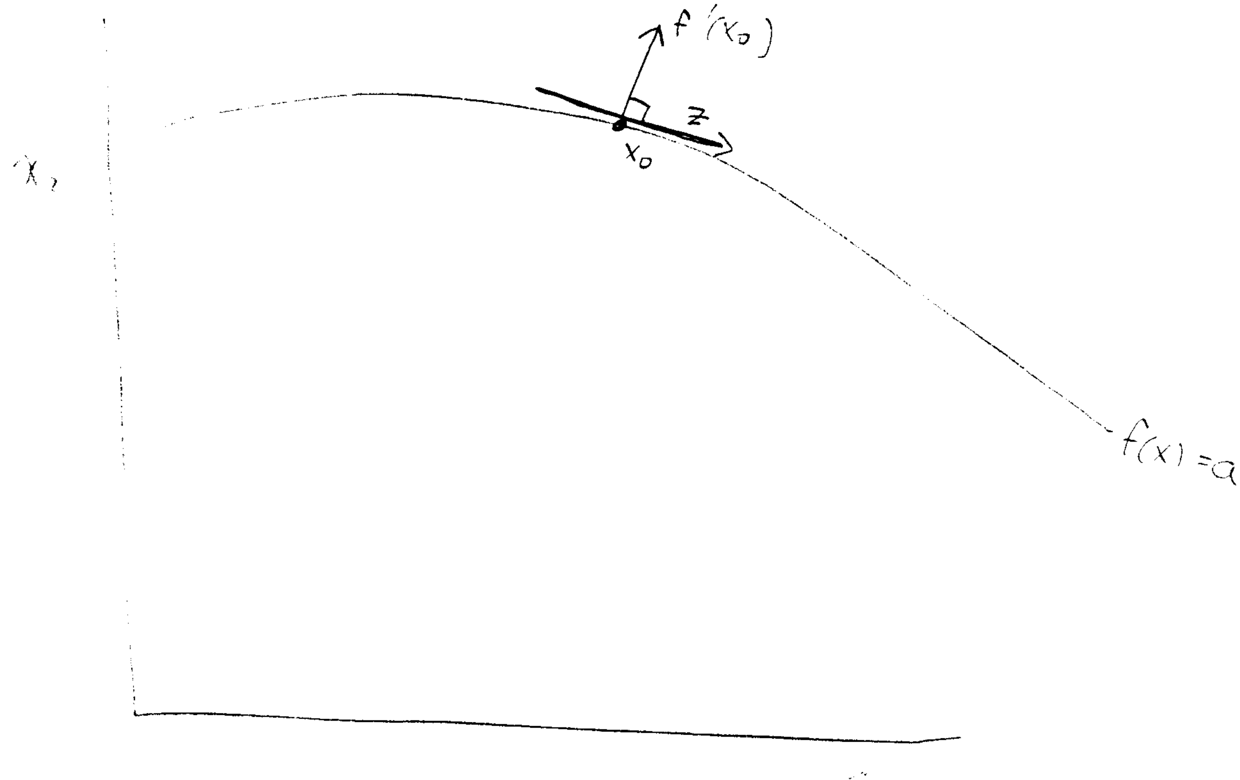

Figure 7.4: The tangent to the level curve $f(x) = a$ is z satisfying $f'(x) \cdot z = 0$.

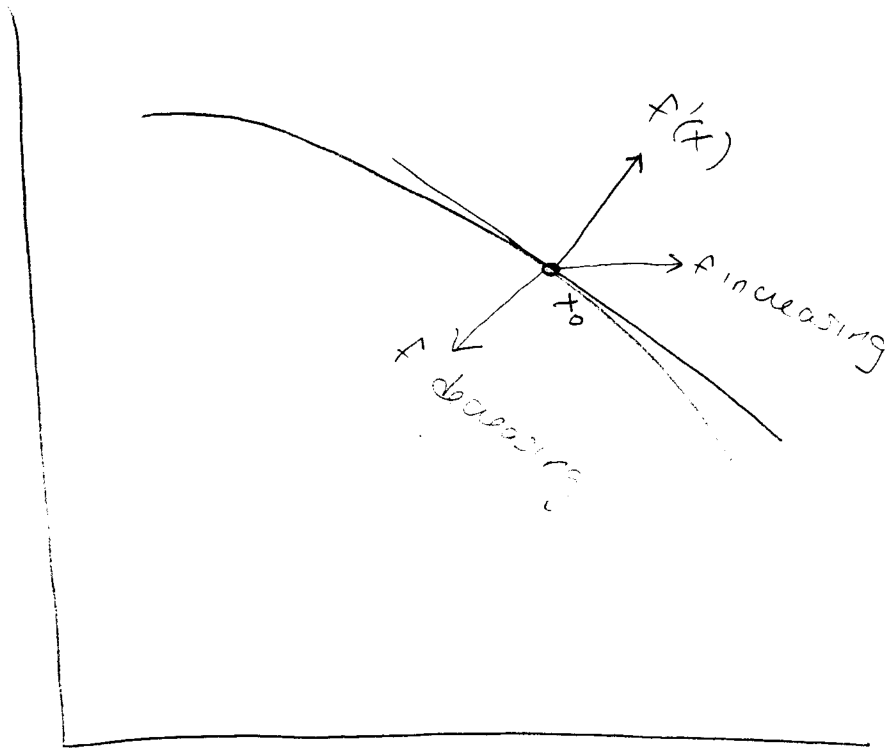

figure 4.5 : f is increasing as we go from X_0 to $X_0 + \lambda Z$ if Z is within 90° of $f'(x_0)$.

summarize the local change in f if we head from x in the direction y, either up ($f'(x)\cdot y > 0$), down ($f'(x)\cdot y < 0$) or the same ($f'(x)\cdot y = 0$).

By Theorem 4.21, concavity of f is equivalent to:

$$f(x) \leq f(y) + f'(y)\cdot(x-y)$$

Suppose $\lambda z = x-y$ is orthogonal to $f'(y)$. Then concavity asserts

$$f(y + \lambda z) \leq f(y) + f'(y)\cdot(\lambda z) = f(y)$$

That is, movement from y in the z direction reduces f (figure 4.6). Another way of putting this is that $\{x/f(x) \geq f(y)\}$ lies totally on one side of the hyperplant defined by

$$\{y + z/f'(y)\cdot z = 0\}.$$

Note, from (4.48) and (4.18) that concavity is also equivalent to:

$f'(x + \lambda y)\cdot y$ decreases in λ

Since $f'(x + \lambda y)\cdot y$ is the change in f as we move from x in the direction y, concavity forces this change to decrease as we move further away (λ increases). Now suppose $f'(x)\cdot z = 0$. Then concavity requires

$$\left.\frac{\partial}{\partial\lambda} f'(x + \lambda z)\cdot z\right|_{\lambda=0} \leq 0$$

That is, the angle formed between $f'(x + \lambda z)$ and z is greater than 90^{o} for $\lambda > 0$. This is illustrated in figure 4.7.

4.4 Odds and Ends

In this section, we present a grab bag of results useful in later chapters.

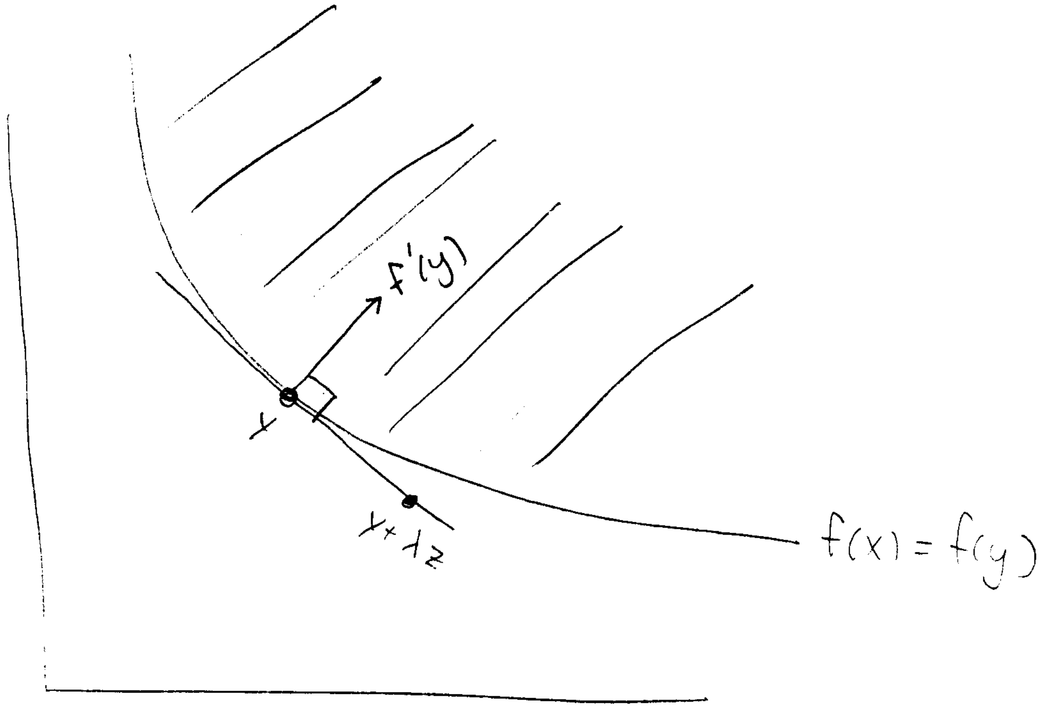

figure 4.6: $\{x \mid f(x) \geq f(y)\}$ (the shaded region) lies totally on one side of $y + \lambda z$ where $f'(y) \cdot z = 0$.

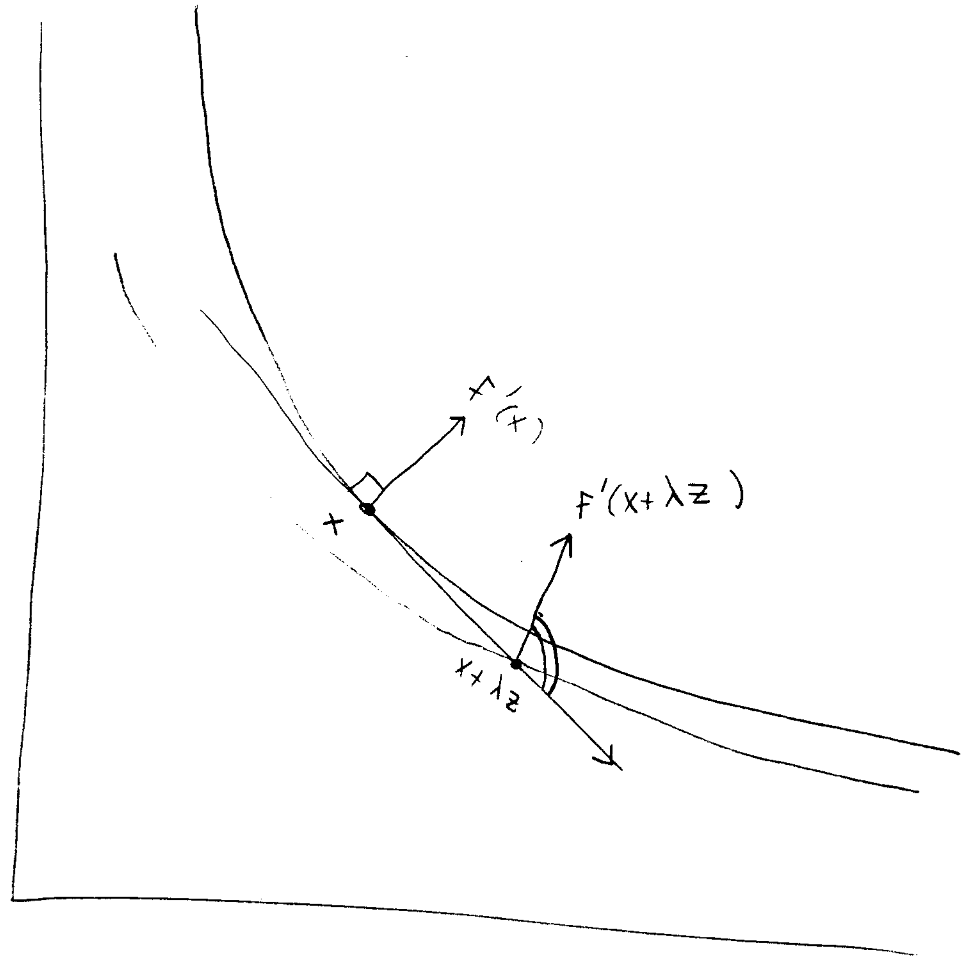

Figure 4.7: The angle formed between $f'(x+\lambda z)$ and $x+\lambda z$ exceeds 90°

If f is concave $f'(x) \cdot z = 0$.

LEMMA 4.24: $x \cdot y$ is a continuous function of x,y.

Proof: Let $\varepsilon > 0$ and fix x_0, y_0. Let

$$\delta = \min \{1, \frac{\varepsilon/3}{1+\|x_0\|}, \frac{\varepsilon/3}{1+\|y_0\|}\}$$

Then, if $\|x-x_0\| < \delta$ and $\|y-y_0\| < \delta$

$$|x \cdot y - x_0 \cdot y_0| = |(x-x_0)\cdot(y-y_0) + x_0\cdot(y-y_0) + y_0\cdot(x-x_0)| \leq$$

$$|(x-x_0)\cdot(y-y_0)| + |x_0\cdot(y-y_0)| + |y_0\cdot(x-x_0)| \leq \qquad \text{(cauchy-schwarz)}$$

$$\|x-x_0\| \, \|y-y_0\| + \|x_0\| \, \|y-y_0\| + \|y_0\| \, \|x-x_0\| <$$

$$1 \left(\frac{\varepsilon/3}{1+\|x_0\|}\right) + \frac{\|x_0\| \; \varepsilon/3}{1 + \|x_0\|} + \frac{\|y_0\| \; \varepsilon/3}{1 + \|y_0\|} \leq \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.$$

Q.E.D.

DEFINITION 4.13: $A \subseteq R^n$ is compact if, for every sequence $x_n \in A$, there is a subsequence $x_{j(n)}$ converging to some $x_0 \in A$. Subsequences delete terms of the original sequence: if $j(1) < j(2) < j(3) < \ldots$, then $x_{j(1)}, x_{j(2)}, \ldots$ is a subsequence.

EXAMPLE 4.19: Let $x_n = (-1)^n$, the alternating sequence $-1, 1, -1, \ldots$. Then one subsequence which converges is every other term: $j(n) = 2n$, so the subsequence $x_2, x_4, x_6, \ldots$ converges (since it is constant at 1).

THEOREM 4.25: $A \subseteq R^n$ is compact if and only if A is closed and bounded.

Proof: ($\Rightarrow$) By contrapositive, we show if A is not closed or not bounded, then A is not compact. If A is not closed, there is a convergent sequence x_n converging to $x_0 \notin A$. But this implies every subsequence of x_n converges to $x_0 \notin A$, and thus A is not compact (see Exercise 25). If A is not bounded, then for each $n \in N$, $\exists \; x_n \in A$, $\|x_n\| \geq n$. This defines a sequence which has no convergent subsequences (exercise 26) and thus A is not compact.

($\Leftarrow$) Suppose A is closed and bounded. Let x_n be a sequence in A. By an R^n version of Bolzano-Weierstrauss (proof is similar to proof of theorem 2.7) there is an x_0 so that, for any $\varepsilon > 0$, infinitely many members of the $\{x_n/n=1,2,...\}$ satisfy $\|x_n - x_0\| < \varepsilon$. But, then, this means for any natural number m, we may choose an $n > m$ with $\|x_n - x_0\| < \varepsilon$. Define the subsequence as follows. Let $x_{j(1)} = x_1$. Given $x_{j(k)}$, choose an n so that n $> j(k)$ and

$\|x_n - x_0\| < 1/k+1$. This $n = j(k+1)$. Clearly the subsequence $x_{j(k)}$ converges to x_0, and since A is closed, $x_0 \in A$.

Q.E.D.

THEOREM 4.26: Suppose $f:R^n \to R$ is continuous and A is compact. Then $\exists\ x^* \in A, (\forall x \in A)\ f(x^*) \geq f(x)$.

Proof: Let $\alpha = \sup\{f(x)/x \in A\}$. Then, by Theorem 2.11, there is a sequence of $f(x_n) \to \alpha$, with $x_n \in A$. Since A is compact, there is a convergent subsequence $x_{j(n)} \to x^* \in A$. Thus, by continuity of f:

$$f(x^*) = f(\lim_{n\to\infty} x_{j(n)}) = \lim_{n\to\infty} f(x_{j(n)}) = \alpha.$$

Q.E.D

DEFINITION 4.14: $f:R^n \to R^n$ is a contraction mapping if $\exists\lambda$, $0 \leq \lambda < 1$, $\forall x, y \in R^n$

$$\|f(x) - f(y)\| < \lambda\|x-y\|$$

THEOREM 4.27: If $f:R^n \to R^n$ is a contraction mapping, then $\exists! x^* \in R^n$, $f(x^*) = x^*$.

Proof: The proof is left as an exercise, with the hint to examin the proof of theorem 2.26.

EXERCISES

1. Show $f(x_1,x_2) = x_1^2 + 2x_1x_2 + x_2^2$ is continuous.

2. Show that $f_i(x) = x_i$, $x \in R^n$, is continuous.

3. If $f(x) = Ax$ for mXn matrix A, show f is continuous.

4. If $f:R^n \to R$ is given by $f(x) = x^TAs$ for nXn matrix A, show f is continuous.

5. Prove Theorem 4.2.

6. Prove Theorem 4.6.

7. Show that any interval $(a,b) = \{x/a < x < b\}$ is open, and any interval $[a,b] = \{x/a \leq x \leq b\}$ is closed.

8. Show, in example 4.4, $\bigcap_{i \in \Gamma} A_i = \{0\}$. Prove $\{0\}$ is not open.

9. Prove any finite set of real numbers $\{a_1,\ldots,a_n\}$ is closed and not open.

10. Prove that the only clopen sets (both open and closed) in R^n are 0 and R^n. That is, show that if $A \subseteq R$ is open and closed, A = 0 or A = R. Hint: use the contrapositive.

11. Prove Theorem 4.6.

12. Show by example that if $f:R^n \to R^m$ is continuous and B is closed, $f^{-1}(B)$ may not be closed. (Hinte: Let n = m = 1 and draw a picture of the function).

13. Show the 2X2 matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is invertible if and only if $a_{11}a_{22} - a_{12}a_{21} \neq 0$.

14. Prove (4.11)-(4.14).

15. Prove, if A and B are nXn symmetric matrices that AB is symmetric.

16. For general 2X2 matrices, characterize the conditions making them positive semidefinite. Hint: note $x^T Ax = a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2$. When is this always at least zero?

17. Show that, if A is an nXn symmetric matrix, and P is the matrix whose columns are eigenvectors of A, then $P^{-1}AP$ is a diagonal matrix: all the elements of the diagonal are zero, and the diagonal elements are the eigenvalues of A.

18. Find the partial derivatives of

$$f(x_1,x_2) = x_1^2 + 2x_1x_2 + x_2^2$$

Find the level curves of this function.

19. Carefully prove that the function in Example 4.15 is discontinuous at (0,0), but all directional derivatives exist at (0,0).

20. If $f:R^n \to R$ is given by $f(x) = x^T Ax$ for nXn matrix x, show

$$f'(x) = Ax + A^T x$$

Show $f''(x) = A + A^T$, and that it is symmetric.

21. Prove Theorem 4.16 directly, without using Theorem 4.14.

22. Prove Theorem 4.17.

23. Prove Theorem 4.21.

24. Prove Theorem 4.27.

25. Show that if a sequence x_n converges to x_0, then every subsequence $x_{j(n)}$ converges to x_0.

26. Show that, if, for all n, $\|x_n\| > n$, no subsequence of x_n converges.