

MATHEMATICS 23a/E-23a, Fall 2018

Linear Algebra and Real Analysis I

Week 11 (Differentiability, Newton's method, inverse and implicit functions,
manifolds)

Author: Paul Bamberg

R scripts by Paul Bamberg

Last modified: November 28, 2018 by Paul Bamberg (fixed reference in workshop
problem 2b)

Reading

- Hubbard, section 1.7 (you have already read most of this)
- Hubbard, sections 1.8 and 1.9 (computing derivatives and differentiability)
- Hubbard, section 2.8 page 233-235 and page 246. (Newton's method)
- Hubbard, section 2.10 up through page 264. (inverse function theorem)
- Hubbard, Section 3.1 (Implicit functions and manifolds)

Recorded Lectures

- Lecture 22 (Week 11, Class 1) (watch on November 20 or 21)
- Lecture 23 (Week 11, Class 2) (watch on November 27 or 28)
- Lecture 24 (Fortnight 12, Class 1) (watch on November 29 or 30)

Proofs to present in section or to a classmate who has done them.

- 11.1 Let $U \subset \mathbb{R}^n$ be an open set, and let f and g be functions from U to \mathbb{R} . Prove that if f and g are differentiable at \mathbf{a} then so is fg , and that

$$[\mathbf{D}(fg)(\mathbf{a})] = f(\mathbf{a})[\mathbf{D}g(\mathbf{a})] + g(\mathbf{a})[\mathbf{D}f(\mathbf{a})].$$

- 11.2 Using the mean value theorem, prove that if a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ has partial derivatives D_1f and D_2f that are continuous at \mathbf{a} , it is differentiable at \mathbf{a} and its derivative is the Jacobian matrix $[D_1f(\mathbf{a}) \ D_2f(\mathbf{a})]$.
- 11.3 The implicit function theorem

Let W be an open subset of \mathbb{R}^n , and let $\mathbf{F} : W \rightarrow \mathbb{R}^{n-k}$ be a C^1 mapping such that $\mathbf{F}(\mathbf{c}) = \mathbf{0}$. Assume that $[\mathbf{D}\mathbf{F}(\mathbf{c})]$ is onto.

Prove that the n variables can be ordered so that the first $n - k$ columns of $[\mathbf{D}\mathbf{F}(\mathbf{c})]$ are linearly independent, and that $[\mathbf{D}\mathbf{F}(\mathbf{c})] = [A|B]$ where A is an invertible $(n - k) \times (n - k)$ matrix.

Set $\mathbf{c} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}$, where \mathbf{a} are the $n - k$ passive variables and \mathbf{b} are the k active variables.

Let \mathbf{g} be the “implicit function” from a neighborhood of \mathbf{b} to a neighborhood of \mathbf{a} such that $\mathbf{g}(\mathbf{b}) = \mathbf{a}$ and $\mathbf{F} \begin{pmatrix} \mathbf{g}(\mathbf{y}) \\ \mathbf{y} \end{pmatrix} = \mathbf{0}$.

Prove that $[\mathbf{D}\mathbf{g}(\mathbf{b})] = -A^{-1}B$.

R Scripts

- Script 3.3A-ComputingDerivatives.R
 - Topic 1 - Testing for differentiability
 - Topic 2 - Illustrating the derivative rules
- Script 3.3B-NewtonsMethod.R
 - Topic 1 - Single variable
 - Topic 2 - 2 equations, 2 unknowns
 - Topic 3 - Three equations in three unknowns
- Script 3.3C-InverseFunction.R
 - Topic 1 - A parametrization function and its inverse
 - Topic 2 - Visualizing coordinates by means of a contour plot
 - Topic 3 - An example that is economic, not geometric

1 Executive Summary

1.1 Definition of the derivative

- Converting the derivative to a matrix

The linear function $f(h) = mh$ is represented by the 1×1 matrix $[m]$.

When we say that $f'(a) = m$, what we mean is that the function

$f(a + h) - f(a)$ is well approximated, for small h , by the linear function mh . The error made by using the approximation is a “remainder” $r(h) = f(a + h) - f(a) - mh$. If f is differentiable, this remainder approaches 0 faster than h , i.e.

$$\lim_{h \rightarrow 0} \frac{r(h)}{h} = \lim_{h \rightarrow 0} \frac{f(a + h) - f(a) - mh}{h} = 0.$$

This definition leads to the standard rule for calculating the number m ,

$$m = \lim_{h \rightarrow 0} \frac{f(a + h) - f(a)}{h}.$$

- Extending this definition to $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$

A linear function $L(\vec{\mathbf{h}})$ is represented by an $m \times n$ matrix.

When we say that \mathbf{f} is differentiable at \mathbf{a} , we mean that the function

$\mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a})$ is well approximated, for any $\vec{\mathbf{h}}$ whose length is small, by a linear function L , called the derivative $[\mathbf{Df}(\mathbf{a})]$.

The error made by using the approximation is a “remainder”

$$\mathbf{r}(\vec{\mathbf{h}}) = \mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a}) - [\mathbf{Df}(\mathbf{a})](\vec{\mathbf{h}}).$$

\mathbf{f} is called differentiable if this remainder approaches 0 faster than $|\vec{\mathbf{h}}|$, i.e.

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}} \frac{1}{|\vec{\mathbf{h}}|} \mathbf{r}(\vec{\mathbf{h}}) = \lim_{\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a}) - [\mathbf{Df}(\mathbf{a})](\vec{\mathbf{h}})) = \mathbf{0}.$$

In that case, $[\mathbf{Df}(\mathbf{a})]$ is represented by the Jacobian matrix $[\mathbf{Jf}(\mathbf{a})]$.

Proof: Since L exists and is linear, it is sufficient to consider its action on each standard basis vector. We choose $\vec{\mathbf{h}} = t\vec{\mathbf{e}}_i$ so that $|\vec{\mathbf{h}}| = t$. Knowing that the limit exists, we can use any sequence that converges to the origin to evaluate it, and so

$$\lim_{t \rightarrow 0} \frac{1}{t} (\mathbf{f}(\mathbf{a} + t\vec{\mathbf{e}}_i) - \mathbf{f}(\mathbf{a}) - tL(\vec{\mathbf{e}}_i)) = 0? \text{ and } L(\vec{\mathbf{e}}_i) = \lim_{t \rightarrow 0} \frac{1}{t} (\mathbf{f}(\mathbf{a} + t\vec{\mathbf{e}}_i) - \mathbf{f}(\mathbf{a}))$$

What is hard is proving that f is differentiable – that L exists – since that requires evaluating a limit where $\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}$. Eventually we will prove that f is differentiable at \mathbf{a} if all its partial derivatives are continuous there.

1.2 Proving differentiability and calculating derivatives

In every case \mathbf{f} is a function from U to \mathbb{R}^m , where U is an open subset of \mathbb{R}^n .

- \mathbf{f} is constant: $\mathbf{f} = \mathbf{c}$. Then $[\mathbf{Df}(\mathbf{a})]$ is the zero linear transformation, since

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a}) - [\mathbf{Df}(\mathbf{a})]\vec{\mathbf{h}}) = \lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{c} - \mathbf{c} - \vec{0}) = \vec{0}.$$

- \mathbf{f} is affine: a constant plus a linear function, $\mathbf{f} = \mathbf{c} + L$. $[\mathbf{Df}(\mathbf{a})] = L$, since

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a}) - [\mathbf{Df}(\mathbf{a})]\vec{\mathbf{h}}) = \lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{c} + L(\mathbf{a} + \vec{\mathbf{h}}) - (\mathbf{c} + L(\mathbf{a})) - L(\vec{\mathbf{h}})) = 0.$$

\mathbf{f} has differentiable components: if $\mathbf{f} = \begin{pmatrix} f_1 \\ \vdots \\ \vdots \\ \vdots \\ f_n \end{pmatrix}$: then $\mathbf{Df}(\mathbf{a}) = \begin{bmatrix} \mathbf{D}f_1(\mathbf{a}) \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{D}f_n(\mathbf{a}) \end{bmatrix}$

- $\mathbf{f} + \mathbf{g}$ is the sum of two functions \mathbf{f} and \mathbf{g} , both differentiable at \mathbf{a} .
The derivative of $\mathbf{f} + \mathbf{g}$ is the sum of the derivatives of \mathbf{f} and \mathbf{g} . (easy to prove)
- $f\mathbf{g}$ is the product of scalar-valued function f and vector-valued \mathbf{g} , both differentiable. Then
 $[\mathbf{D}(f\mathbf{g})(\mathbf{a})]\vec{\mathbf{v}} = f(\mathbf{a})([\mathbf{Dg}(\mathbf{a})]\vec{\mathbf{v}}) + ([\mathbf{Df}(\mathbf{a})]\vec{\mathbf{v}})\mathbf{g}(\mathbf{a})$.
- \mathbf{g}/f is the quotient of vector-valued function \mathbf{g} and scalar-valued f , both differentiable, and $f(\mathbf{a}) \neq 0$. Then

$$[\mathbf{D}(\frac{\mathbf{g}}{f})(\mathbf{a})]\vec{\mathbf{v}} = \frac{[\mathbf{Dg}(\mathbf{a})]\vec{\mathbf{v}}}{f(\mathbf{a})} - \frac{([\mathbf{Df}(\mathbf{a})]\vec{\mathbf{v}})\mathbf{g}(\mathbf{a})}{(f(\mathbf{a}))^2}.$$

- $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are open sets, and \mathbf{a} is a point in U at which we want to evaluate a derivative.

$\mathbf{g} : U \rightarrow V$ is differentiable at \mathbf{a} , and $[\mathbf{Dg}(\mathbf{a})]$ is an $m \times n$ Jacobian matrix.

$\mathbf{f} : V \rightarrow \mathbb{R}^p$ is differentiable at $\mathbf{g}(\mathbf{a})$, and $[\mathbf{Df}(\mathbf{g}(\mathbf{a}))]$ is a $p \times m$ Jacobian matrix.

The chain rule states that $[\mathbf{D}(\mathbf{f} \circ \mathbf{g})(\mathbf{a})] = [\mathbf{Df}(\mathbf{g}(\mathbf{a}))] \circ [\mathbf{Dg}(\mathbf{a})]$.

- The combined effect of all these rules is effectively that if a function is defined by well-behaved formulas (no division by zero), it is differentiable, and its derivative is represented by its Jacobian matrix.

1.3 Connection between Jacobian matrix and derivative

- If $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is defined on an open set $U \in \mathbb{R}^n$, and

$$\mathbf{f}(\mathbf{x}) = \mathbf{f} \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} f_1(\mathbf{x}) \\ \dots \\ f_m(\mathbf{x}) \end{pmatrix}$$

the Jacobian matrix $[\mathbf{J}\mathbf{f}(\mathbf{x})]$ is made up of all the partial derivatives of \mathbf{f} :

$$[\mathbf{J}\mathbf{f}(\mathbf{a})] = \begin{bmatrix} D_1 f_1(\mathbf{a}) & \dots & D_n f_1(\mathbf{a}) \\ \dots & \dots & \dots \\ D_1 f_m(\mathbf{a}) & \dots & D_n f_m(\mathbf{a}) \end{bmatrix}$$

- We can invent pathological cases where the Jacobian matrix of f exists (because all the partial derivatives exist), but the function f is not differentiable. In such a case, using the formula

$$\nabla_{\vec{v}} f(\mathbf{a}) = [\mathbf{J}f(\mathbf{a})]\vec{v}$$

generally gives the wrong answer for the directional derivative! You are trying to use a linear approximation where none exists.

- Using the Jacobian matrix of partial derivatives to get a good affine approximation for $f(\mathbf{a} + \vec{\mathbf{h}})$ is tantamount to assuming that you can reach the point $\mathbf{a} + \vec{\mathbf{h}}$ by moving along lines that are parallel to the coordinate axes and that the change in the function value along the solid horizontal line is well approximated by the change along the dotted horizontal line. With the aid of the mean value theorem, you can show that this is the case if (proof 11.2) the partial derivatives of f at \mathbf{a} are continuous.

$$\begin{array}{ccc} (a_1, a_2 + h_2) & (a_1 + h_1, a_2 + h_2) & \\ & \overline{\hspace{1.5cm}} & \\ & \vdots & \\ (a_1, a_2) & (a_1 + h_1, a_2) & \end{array}$$

1.4 Newton's method – one variable

Newton's method is based on the tangent-line approximation. Function f is differentiable. We are trying to solve the equation $f(x) = 0$, and we have found a value a_0 that is close to the desired x . So we use the best affine approximation $f(x) \approx f(a_0) + f'(x_0)(x - a_0)$.

Then we find a value a_1 for which this tangent-line approximation equals zero.

$f(a_0) + f'(x_0)(a_1 - a_0) = 0$, and $a_1 = a_0 - f(a_0)/f'(a_0)$.

When $f(a_0)$ is small, $f'(a_0)$ is large, and $f'(a_0)$ does not change too rapidly, a_1 is a much improved approximation to the desired solution x . Details, for which Kantorovich won the Nobel prize in economics, are in Hubbard.

1.5 Newton's method – more than one variable

Example: we are trying to solve a system of n nonlinear equations in n unknowns, e.g.

$$x^2 e^y - \sin(y) - 0.3 = 0$$

$$\tan x + x^2 y^2 - 1 = 0.$$

Ordinary algebra is no help – there is no nonlinear counterpart to row reduction. U is an open subset of \mathbb{R}^n , and we have a differentiable function $\vec{f}(\mathbf{x}) : U \rightarrow \mathbb{R}^n$.

In the example, $\vec{f} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 e^y - \sin(y) - 0.3 \\ \tan x + x^2 y^2 - 1 \end{pmatrix}$, which is differentiable.

We are trying to solve the equation $\vec{f}(\mathbf{x}) = \vec{0}$.

Suppose we have found a value \mathbf{a}_0 that is close to the desired \mathbf{x} .

Again we use the best affine approximation

$$\vec{f}(\mathbf{x}) \approx \vec{f}(\mathbf{a}_0) + [\mathbf{D}\vec{f}(\mathbf{a}_0)](\mathbf{x} - \mathbf{a}_0).$$

We set out to find a value \mathbf{a}_1 for which this affine approximation equals zero.

$$\vec{f}(\mathbf{a}_0) + [\mathbf{D}\vec{f}(\mathbf{a}_0)](\mathbf{a}_1 - \mathbf{a}_0) = \vec{0}$$

This is a linear equation, which we know how to solve!

If $[\mathbf{D}\vec{f}(\mathbf{a}_0)]$ is invertible (and if it is not, we look for a better \mathbf{a}_0), then

$$\mathbf{a}_1 = \mathbf{a}_0 - [\mathbf{D}\vec{f}(\mathbf{a}_0)]^{-1} \vec{f}(\mathbf{a}_0).$$

Iterating this procedure is the best known for solving systems of nonlinear equations. Hubbard has a detailed discussion (which you are free to ignore) of how to use Kantorovich's theorem to assess convergence.

1.6 The inverse function theorem – short version

For function $f : [a, b] \rightarrow [c, d]$, we know that if f is strictly increasing or strictly decreasing on interval $[a, b]$, there is an inverse function g for which $g \circ f$ and $f \circ g$ are both the identity function. We can find $g(y)$ for a specific y by solving $f(x) - y = 0$, perhaps by Newton's method. If $f(x_0) = y_0$ and $f'(x_0) \neq 0$, we can prove that g is differentiable at y_0 and that $g'(y_0) = 1/f'(x_0)$.

“Strictly monotone” does not generalize, but “nonzero $f'(x_0)$ ” generalizes to “invertible $[\mathbf{D}\mathbf{f}(\mathbf{x}_0)]$.” Start with a function $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ whose partial derivatives are all continuous, so that we know that it is differentiable everywhere. Choose a point \mathbf{x}_0 where the derivative $[\mathbf{D}\mathbf{f}(\mathbf{x}_0)]$ is an invertible matrix. Set $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$. Then there is a differentiable local inverse function $\mathbf{g} = \mathbf{f}^{-1}$ such that

- $\mathbf{g}(\mathbf{y}_0) = \mathbf{x}_0$.
- $\mathbf{f}(\mathbf{g}(\mathbf{y})) = \mathbf{y}$ if \mathbf{y} is close enough to \mathbf{y}_0 .
- $[\mathbf{D}\mathbf{g}(\mathbf{y})] = [\mathbf{D}\mathbf{f}(\mathbf{g}(\mathbf{y}))]^{-1}$ (follows from the chain rule)

1.7 Implicit functions – review of the linear case.

We have n unknowns, $n - k$ equations, e.g for $n = 3, k = 1$

$$2x + 3y - z = 0, 4x - 2y + 3z = 0$$

Create an $(n - k) \times n$ matrix: $T = \begin{bmatrix} 2 & 3 & -1 \\ 4 & -2 & 3 \end{bmatrix}$

If the matrix T is not onto, its rows (the equations) are linearly dependent. Otherwise, when we row reduce, we will find $n - k = 2$ pivotal columns and $k = 1$ nonpivotal columns. We assign values arbitrarily to the “active” variables that correspond to the nonpivotal columns, and then the values of the “passive” variables that corresponds to the pivotal column are determined.

Suppose that we reorder the unknowns so that the “active” variables come last. Then, after we row reduce the matrix, the first $n - k$ columns will be pivotal. So the first $n - k$ columns will be linearly independent, and they form an invertible square matrix. The matrix is now of the form $T = [A|B]$, where A is invertible.

The solution vector is of the form $\vec{v} = \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix}$, where the passive variables \vec{x} come first, the active variables \vec{y} come second.

A solution to $T\vec{v} = \vec{0}$ is obtained by choosing \vec{y} arbitrarily and setting $\vec{x} = -A^{-1}B\vec{y}$. Our system of equations determines \vec{x} “implicitly” in terms of \vec{y} .

1.8 Implicit function theorem – the nonlinear case.

We have a point $\mathbf{c} \in \mathbb{R}^n$, a neighborhood W of \mathbf{c} , and a function $\mathbf{F} : W \rightarrow \mathbb{R}^{n-k}$ for which $\mathbf{F}(\mathbf{c}) = 0$ and $[\mathbf{DF}(\mathbf{c})]$ is onto. \mathbf{F} imposes *constraints*.

The variables are ordered so that the $n - k$ pivotal columns in the Jacobian matrix, which correspond to the passive variables, come first. Let \mathbf{a} denote the passive variables at \mathbf{c} ; let \mathbf{b} denote the active variables at \mathbf{c} .

The implicit function \mathbf{g} expresses the passive variables in terms of the active variables, and $\mathbf{g}(\mathbf{b}) = \mathbf{a}$. For \mathbf{y} near \mathbf{b} , $\mathbf{x} = \mathbf{g}(\mathbf{y})$ determines passive variables such that $\mathbf{F} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \mathbf{0}$. Tweak \mathbf{y} , and \mathbf{g} specifies how to tweak \mathbf{x} so that the constraints are still satisfied.

Although we usually cannot find a formula for \mathbf{g} , we can find its derivative at \mathbf{b} by the same recipe that worked in simple cases.

Evaluate the Jacobian matrix $[\mathbf{DF}(\mathbf{c})]$.

Extract the first $n - k$ columns to get an invertible square matrix A .

Let the inverse of this matrix act on the remaining k columns (matrix B) and change the sign to get the $(n - k) \times k$ Jacobian matrix for \mathbf{g} .

That is, $[\mathbf{Dg}(\mathbf{b})] = -A^{-1}B$.

1.9 Curves, Surfaces, Graphs, and Manifolds

Manifolds are a generalization of smooth curves and surfaces.

The simplest sort of manifold is a flat one, described by linear equations. An example is the line of slope 2 that passes through the point $x = 0, y = -2$: a one-dimensional submanifold of \mathbb{R}^2

There are three equivalent ways to describe such a manifold.

- (The definition) As the *graph* of a function that expresses the passive variables in terms of the active variables: either $y = f(x) = -2 + 2x$ or $x = g(y) = \frac{1}{2}(y + 2)$.
- As a “locus” defined by a constraint equation $F\begin{pmatrix} x \\ y \end{pmatrix} = 2x - y - 2 = 0$.
- By a parametrization function $g(t) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + t \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

Definition: A subset $M \subset \mathbb{R}^n$ is a smooth manifold if locally it is the graph of a C^1 function (the partial derivatives are continuous). “Locally” means that for any point $\mathbf{x} \in M$ we can find a neighborhood U of \mathbf{x} such that within $M \cap U$, there is a C^1 function that expresses $n - k$ passive variables in terms of the other k active variables. The number k is the dimension of the manifold. In \mathbb{R}^3 there are four possibilities:

- $k = 3$. Any open subset $M \subset \mathbb{R}^3$ is a smooth 3-dimensional manifold. In this case $k = 3$, and the manifold is the graph of a function $f : \mathbb{R}^3 \rightarrow \{\vec{0}\}$, whose codomain is the trivial vector space $\{\vec{0}\}$ that contains just a single point. Such a function is necessarily constant, and its derivative is zero.
- $k = 2$. The graph of $z = f\begin{pmatrix} x \\ y \end{pmatrix} = x^2 + y^2$ is a paraboloid.
- $k = 1$. The graph of the function $\begin{pmatrix} x \\ y \end{pmatrix} = \vec{f}(z) = \begin{pmatrix} \cos 2\pi z \\ \sin 2\pi z \end{pmatrix}$ is a helix.
- $k = 0$. In this case the manifold consists of one or more isolated points. Near any of these points \mathbf{x}_0 , it is the graph of a function $\vec{f} : \{\vec{0}\} \rightarrow \mathbb{R}^3$ whose domain is a zero-dimensional vector space and whose image is the point $\mathbf{x}_0 \in \mathbb{R}^3$.

There is no requirement that a manifold be the graph of a single function, or that the “active” variables be the same at every point on the manifold. The unit circle, the locus of $x^2 + y^2 - 1 = 0$, is the union of four function graphs, two of which have x as the active variable, two of which have y . By using a parameter t that is not one of the variables, we can represent it by the parametrization $\begin{pmatrix} x \\ y \end{pmatrix} = g(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$

2 Lecture outline

1. The derivative as a linear transformation

When we say that function f is differentiable at $x = a$ and that $f'(a) = m$, what we mean is that the function $f(a + h) - f(a)$ is well approximated, for small h , by a linear function $L(h) = mh$, where $m = f'(a)$.

Show how this idea can be viewed as a “tangent-line approximation” to $f(x)$ for x near to a .

In the single-variable case, we usually think of the derivative $f'(a)$ as just a number, not a linear function of an increment h , but that view will not generalize to derivatives in \mathbb{R}^n . Here is a view of single-variable calculus that generalizes correctly.

Any linear function $L(h) = mh$ is represented by the 1×1 matrix $[m]$, which in turn is represented by the real number m .

The error made by using the tangent-line approximation $f(a + h) - f(a) = f'(a)h$ is a “remainder”

$$r(h) = f(a + h) - f(a) - f'(a)h.$$

If f is differentiable, this remainder approaches 0 faster than h , i.e.

$$\lim_{h \rightarrow 0} \frac{r(h)}{h} = \lim_{h \rightarrow 0} \frac{f(a + h) - f(a) - f'(a)h}{h} = 0.$$

This definition leads to the standard rule for calculating the number $f'(a)$,

$$f'(a) = \lim_{h \rightarrow 0} \frac{f(a + h) - f(a)}{h}.$$

What mathematical object represents a linear transformation $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$?

2. Extending this definition to $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$

A linear function $L(\vec{\mathbf{h}})$ is represented by an $m \times n$ matrix.

What matrix represents the linear function

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = L \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 \\ 3x_1 - x_2 \end{pmatrix}?$$

When we say that \mathbf{f} is *differentiable* at \mathbf{a} , we mean that the function $\mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a})$ is well approximated, for any $\vec{\mathbf{h}}$ whose length is small, by a linear function L , called the derivative $[\mathbf{Df}(\mathbf{a})]$.

The error made by using the approximation is a “remainder”

$$\mathbf{r}(\vec{\mathbf{h}}) = \mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a}) - [\mathbf{Df}(\mathbf{a})](\vec{\mathbf{h}}).$$

\mathbf{f} is called differentiable if this remainder approaches 0 faster than $|\vec{\mathbf{h}}|$, i.e.

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}} \frac{1}{|\vec{\mathbf{h}}|} \mathbf{r}(\vec{\mathbf{h}}) = \lim_{\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a}) - [\mathbf{Df}(\mathbf{a})](\vec{\mathbf{h}})) = \mathbf{0}.$$

In that case, $[\mathbf{Df}(\mathbf{a})]$ is represented by the Jacobian matrix $[\mathbf{Jf}(\mathbf{a})]$.

Proof: Since L exists and is linear, it is sufficient to consider its action on each standard basis vector. We choose $\vec{\mathbf{h}} = t\vec{\mathbf{e}}_i$ so that $|\vec{\mathbf{h}}| = t$. Knowing that the limit exists, we can use any sequence that converges to the origin to evaluate it, and so

$$\lim_{t \rightarrow 0} \frac{1}{t} (\mathbf{f}(\mathbf{a} + t\vec{\mathbf{e}}_i) - \mathbf{f}(\mathbf{a}) - tL(\vec{\mathbf{e}}_i)) = 0? \text{ and } L(\vec{\mathbf{e}}_i) = \lim_{t \rightarrow 0} \frac{1}{t} (\mathbf{f}(\mathbf{a} + t\vec{\mathbf{e}}_i) - \mathbf{f}(\mathbf{a}))$$

What is hard is proving that f is differentiable – that L exists – since that requires evaluating a limit where $\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}$. Such a limit exists only if every sequence $\vec{\mathbf{h}}_1, \vec{\mathbf{h}}_2, \dots$ that converges to $\vec{\mathbf{0}}$ leads to the conclusion that

$$\lim_{\vec{\mathbf{h}}_n \rightarrow \vec{\mathbf{0}}} \frac{1}{|\vec{\mathbf{h}}_n|} \mathbf{r}(\vec{\mathbf{h}}_n) = 0$$

Mere existence of partial derivatives of \mathbf{f} at \mathbf{a} does not guarantee that \mathbf{f} is differentiable at \mathbf{a} . Eventually we will prove (proof 11.2) that f is differentiable at \mathbf{a} if all its partial derivatives are *continuous* there.

3. Proving differentiability and calculating derivatives

In every case \mathbf{f} is a function from U to \mathbb{R}^m , where U is an open subset of \mathbb{R}^n .

- \mathbf{f} is constant: $\mathbf{f} = \mathbf{c}$. Then $[\mathbf{Df}(\mathbf{a})]$ is the zero linear transformation, since

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a}) - [\mathbf{Df}(\mathbf{a})]\vec{\mathbf{h}}) = \lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{c} - \mathbf{c} - \vec{0}) = \vec{0}.$$

- \mathbf{f} is affine: a constant plus a linear function, $\mathbf{f} = \mathbf{c} + L$. $[\mathbf{Df}(\mathbf{a})] = L$, since

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a}) - [\mathbf{Df}(\mathbf{a})]\vec{\mathbf{h}}) = \lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{c} + L(\mathbf{a} + \vec{\mathbf{h}}) - (\mathbf{c} + L(\mathbf{a})) - L(\vec{\mathbf{h}})) = 0.$$

$$\mathbf{f} \text{ has differentiable components: if } \mathbf{f} = \begin{pmatrix} f_1 \\ \cdot \\ \cdot \\ \cdot \\ f_n \end{pmatrix} : \text{ then } \mathbf{Df}(\mathbf{a}) = \begin{bmatrix} \mathbf{D}f_1(\mathbf{a}) \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{D}f_n(\mathbf{a}) \end{bmatrix}$$

- $\mathbf{f} + \mathbf{g}$ is the sum of two functions \mathbf{f} and \mathbf{g} , both differentiable at \mathbf{a} .
The derivative of $\mathbf{f} + \mathbf{g}$ is the sum of the derivatives of \mathbf{f} and \mathbf{g} . (proof on next page)

4. Derivative of a sum

\mathbf{f} and \mathbf{g} are differentiable at \mathbf{a} . The obvious guess is that the derivative of $\mathbf{f} + \mathbf{g}$ is the sum of their derivatives.

Since \mathbf{f} and \mathbf{g} are differentiable at \mathbf{a} , we know that

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{f}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{f}(\mathbf{a}) - ([\mathbf{D}\mathbf{f}(\mathbf{a})]\vec{\mathbf{h}})) = 0.$$

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} (\mathbf{g}(\mathbf{a} + \vec{\mathbf{h}}) - \mathbf{g}(\mathbf{a}) - ([\mathbf{D}\mathbf{g}(\mathbf{a})]\vec{\mathbf{h}})) = 0.$$

To prove the obvious guess, show that

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{0}} \frac{1}{|\vec{\mathbf{h}}|} ((\mathbf{f} + \mathbf{g})(\mathbf{a} + \vec{\mathbf{h}}) - (\mathbf{f} + \mathbf{g})(\mathbf{a}) - ([\mathbf{D}\mathbf{f}(\mathbf{a})] + [\mathbf{D}\mathbf{g}(\mathbf{a})])\vec{\mathbf{h}})) = 0.$$

5. Product rule (your proof 11.1):

Now comes something harder: the product rule for two scalar-valued functions f and g . It is easy to guess what the derivative of fg must be, since in single variable calculus, $(fg)' = fg' + gf'$.

Product rule: $[\mathbf{D}(fg)(\mathbf{a})] = f(\mathbf{a})[\mathbf{D}g(\mathbf{a})] + g(\mathbf{a})[\mathbf{D}f(\mathbf{a})]$.

- Step 1: Write the “remainder” $r(\vec{\mathbf{h}})$ that must have the property

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}} \frac{r(\vec{\mathbf{h}})}{|\vec{\mathbf{h}}|} = 0.$$

$$r(\vec{\mathbf{h}}) = f(\mathbf{a} + \vec{\mathbf{h}})g(\mathbf{a} + \vec{\mathbf{h}}) - f(\mathbf{a})g(\mathbf{a}) - f(\mathbf{a})[\mathbf{D}g(\mathbf{a})]\vec{\mathbf{h}} - g(\mathbf{a})[\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}}$$

- Step 2 – a trick that must be memorized: Subtract and add $f(\mathbf{a})g(\mathbf{a} + \vec{\mathbf{h}})$, and subtract and add $g(\mathbf{a} + \vec{\mathbf{h}})[\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}}$.

$$r(\vec{\mathbf{h}}) = f(\mathbf{a} + \vec{\mathbf{h}})g(\mathbf{a} + \vec{\mathbf{h}}) - f(\mathbf{a})g(\mathbf{a} + \vec{\mathbf{h}}) + f(\mathbf{a})g(\mathbf{a} + \vec{\mathbf{h}}) - f(\mathbf{a})g(\mathbf{a}) - f(\mathbf{a})[\mathbf{D}g(\mathbf{a})]\vec{\mathbf{h}} - g(\mathbf{a} + \vec{\mathbf{h}})[\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}} + g(\mathbf{a} + \vec{\mathbf{h}})[\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}} - g(\mathbf{a})[\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}}.$$

- Step 3: split into three terms, one involving the remainder for f , one involving the remainder for g , and one involving $[\mathbf{D}f(\mathbf{a})]$.

$$r(\vec{\mathbf{h}}) = r_1(\vec{\mathbf{h}}) + r_2(\vec{\mathbf{h}}) + r_3(\vec{\mathbf{h}}), \text{ where}$$

$$r_1(\vec{\mathbf{h}}) = f(\mathbf{a} + \vec{\mathbf{h}})g(\mathbf{a} + \vec{\mathbf{h}}) - f(\mathbf{a})g(\mathbf{a} + \vec{\mathbf{h}}) - g(\mathbf{a} + \vec{\mathbf{h}})[\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}}$$

$$r_2(\vec{\mathbf{h}}) = f(\mathbf{a})g(\mathbf{a} + \vec{\mathbf{h}}) - f(\mathbf{a})g(\mathbf{a}) - f(\mathbf{a})[\mathbf{D}g(\mathbf{a})]\vec{\mathbf{h}}.$$

$$r_3(\vec{\mathbf{h}}) = g(\mathbf{a} + \vec{\mathbf{h}})[\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}} - g(\mathbf{a})[\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}}.$$

- Step 4: Divide each term by $|\vec{\mathbf{h}}|$, and use the differentiability of f and g to prove that the limit $\lim_{\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}} \frac{r(\vec{\mathbf{h}})}{|\vec{\mathbf{h}}|}$ is zero. For each term you have the product of two factors: one approaches zero, while the other is bounded.

$$r_1(\vec{\mathbf{h}}) = (f(\mathbf{a} + \vec{\mathbf{h}}) - f(\mathbf{a}) - [\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}})g(\mathbf{a} + \vec{\mathbf{h}}).$$

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}} \frac{r_1(\vec{\mathbf{h}})}{|\vec{\mathbf{h}}|} = 0,$$

since the first factor over $|\vec{\mathbf{h}}|$ goes to zero and the second is bounded.

$$r_2(\vec{\mathbf{h}}) = f(\mathbf{a})(g(\mathbf{a} + \vec{\mathbf{h}}) - g(\mathbf{a}) - [\mathbf{D}g(\mathbf{a})]\vec{\mathbf{h}}).$$

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}} \frac{r_2(\vec{\mathbf{h}})}{|\vec{\mathbf{h}}|} = 0,$$

since the second factor over $|\vec{\mathbf{h}}|$ goes to zero and the first is constant.

$$r_3(\vec{\mathbf{h}}) = [g(\mathbf{a} + \vec{\mathbf{h}}) - g(\mathbf{a})][\mathbf{D}f(\mathbf{a})]\vec{\mathbf{h}}.$$

$$\lim_{\vec{\mathbf{h}} \rightarrow \vec{\mathbf{0}}} \frac{r_3(\vec{\mathbf{h}})}{|\vec{\mathbf{h}}|} = 0,$$

since the first factor goes to zero by continuity and the second factor over $|\vec{\mathbf{h}}|$ is bounded.

6. (Proof 11.1)

Let $U \subset \mathbb{R}^n$ be an open set, and let f and g be functions from U to \mathbb{R} . Prove that if f and g are differentiable at \mathbf{a} then so is fg , and that

$$[\mathbf{D}(fg)(\mathbf{a})] = f(\mathbf{a})[\mathbf{D}g(\mathbf{a})] + g(\mathbf{a})[\mathbf{D}f(\mathbf{a})].$$

(Simpler than in Hubbard because both f and g are scalar-valued functions)

7. Chain rule in \mathbb{R}^n – not a proof, but still pretty convincing

The chain rule for differentiating composition of functions in general is as simple as you could hope for on the basis of single-variable calculus. Remember the single-variable version:

$$(f \circ g)'(a) = f'(g(a))g'(a).$$

This can be rewritten

$$[D(f \circ g)(a)] = [Df(g(a))][Dg(a)],$$

where the square brackets convert the old-style derivatives (numbers) into 1×1 Jacobian matrices.

This says that the derivative of the composition of f and g is the composition of the linear function “multiply by $g'(a)$ ” and the linear function “multiply by $f'(g(a))$ ” Notice that f has to be differentiable at $g(a)$.

Here is the generalization:

$U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are open sets, and \mathbf{a} is a point in U at which we want to evaluate a derivative.

$\mathbf{g} : U \rightarrow V$ is differentiable at \mathbf{a} , and $[\mathbf{Dg}(\mathbf{a})]$ is a $m \times n$ Jacobian matrix.

$\mathbf{f} : V \rightarrow \mathbb{R}^p$ is differentiable at $\mathbf{g}(\mathbf{a})$, and $[\mathbf{Df}(\mathbf{g}(\mathbf{a}))]$ is a $p \times m$ Jacobian matrix.

The chain rule states that $[\mathbf{D}(\mathbf{f} \circ \mathbf{g})(\mathbf{a})] = [\mathbf{Df}(\mathbf{g}(\mathbf{a}))] \circ [\mathbf{Dg}(\mathbf{a})]$.

Draw a diagram to illustrate what happens in the case $n = m = p = 2$ when you use derivatives to find a linear approximation to

$(\mathbf{f} \circ \mathbf{g})(\mathbf{a} + \tilde{\mathbf{h}}) - (\mathbf{f} \circ \mathbf{g})(\mathbf{a})$.

This approximation can be done in a single step or in two steps.

8. Two easy chain rule examples

- (a) $\mathbf{g} : \mathbb{R} \rightarrow \mathbb{R}^2$ maps time into the position of a particle moving around the unit circle:

$$\mathbf{g}(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}.$$

$f : \mathbb{R}^2 \rightarrow \mathbb{R}$ maps a point into the temperature at that point.

$$f \begin{pmatrix} x \\ y \end{pmatrix} = x^2 - y^2$$

The composition $f \circ \mathbf{g}$ maps time directly into temperature .

Confirm that $[D(f \circ g)(t)] = [\mathbf{D}f(\mathbf{g}(t))] \circ [\mathbf{D}\mathbf{g}(t)]$.

- (b) Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$ be any differentiable function. You can make a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ that is constant on any circle centered at the origin by forming the composition $f \begin{pmatrix} x \\ y \end{pmatrix} = \phi(x^2 + y^2)$.

Show that f satisfies the partial differential equation $yD_1f - xD_2f = 0$.

9. Connection between Jacobian matrix and derivative

- If $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is defined on an open set $U \in \mathbb{R}^n$, and

$$\mathbf{f}(\mathbf{x}) = \mathbf{f} \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} f_1(\mathbf{x}) \\ \dots \\ f_m(\mathbf{x}) \end{pmatrix}$$

the Jacobian matrix $[\mathbf{Jf}(\mathbf{x})]$ is made up of all the partial derivatives of \mathbf{f} :

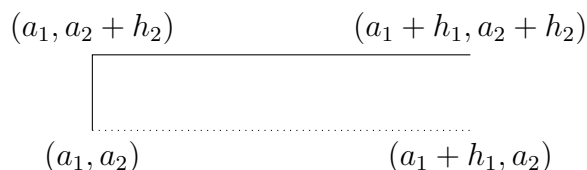
$$[\mathbf{Jf}(\mathbf{a})] = \begin{bmatrix} D_1 f_1(\mathbf{a}) & \dots & D_n f_1(\mathbf{a}) \\ \dots & & \dots \\ D_1 f_m(\mathbf{a}) & \dots & D_n f_m(\mathbf{a}) \end{bmatrix}$$

- We can invent pathological cases where the Jacobian matrix of f exists (because all the partial derivatives exist), but the function f is not differentiable. In such a case, using the formula

$$\nabla_{\vec{v}} f(\mathbf{a}) = [\mathbf{Jf}(\mathbf{a})]\vec{v}$$

generally gives the wrong answer for the directional derivative! You are trying to use a linear approximation where none exists.

- Using the Jacobian matrix of partial derivatives to get a good affine approximation for $f(\mathbf{a} + \vec{\mathbf{h}})$ is tantamount to assuming that you can reach the point $\mathbf{a} + \vec{\mathbf{h}}$ by moving along lines that are parallel to the coordinate axes and that the change in the function value along the solid horizontal line is well approximated by the change along the dotted horizontal line. With the aid of the mean value theorem, you can show that this is the case if (proof 11.2) the partial derivatives of f at \mathbf{a} are continuous.



10. Jacobian matrix for a parametrization function gives a good affine approximation

Here is the function that converts the latitude u and longitude v of a point on the unit sphere to the Cartesian coordinates of that point.

$$f \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos u \cos v \\ \cos u \sin v \\ \sin u \end{pmatrix}$$

Work out the Cartesian coordinates of the point with $\sin u = \frac{3}{5}$ (37 degrees North latitude) and $\sin v = 1$ (90 degrees East longitude), and calculate the Jacobian matrix at that point. Then find the best affine approximation to the Cartesian coordinates of the nearby point where u is 0.01 radians less (going south) and v is 0.02 radians greater (going east).

11. A non-differentiable function

Consider a surface where the height z is given by the function

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \frac{3x^2y - y^3}{x^2 + y^2}; f\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0.$$

This function is not differentiable at the origin, and so you cannot calculate its directional derivatives there by using the Jacobian matrix!

- (a) Along the first standard basis vector, the directional derivative at the origin is zero. Find two vectors along other directions that also have this property.
- (b) Along the second standard basis vector, the directional derivative at the origin is -1.
Find two vectors along other directions that also have this property. (This surface is sometimes called a “monkey saddle,” because a monkey could sit comfortably on it with its two legs and its tail placed along these three downward-sloping directions.)
- (c) Calculate the directional derivative along an arbitrary unit vector $\vec{e}_\theta = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$. Using the trig identity $\sin 3\theta = 3 \sin \theta \cos^2 \theta - \sin^3 \theta$, quickly rederive the special cases of parts (a) and (b).
- (d) Using the definition of the derivative, give a convincing argument that this function is not differentiable at the origin.

12. The mean-value theorem in \mathbb{R}^n

For functions of one variable, this is an old friend.

If $f : [a, b] \rightarrow \mathbb{R}$ is continuous on $[a, b]$ and differentiable on (a, b) , then $\exists c \in [a, b]$ such that

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

The generalization uses the segment from \mathbf{a} to \mathbf{b} like the closed interval $[a, b]$.

The function f (takes values in \mathbb{R}) must be differentiable on an open set U that includes this entire segment.

The conclusion is that

$$f(\mathbf{b}) - f(\mathbf{a}) = [\mathbf{D}f(\mathbf{c})](\vec{\mathbf{b} - \mathbf{a}}).$$

The proof (Hubbard p. 148) is easy. Define a function $\mathbf{h}(t)$ that maps the interval $0 \leq t \leq 1$ uniformly into the segment from \mathbf{a} to \mathbf{b} . The formula is

$$\mathbf{h}(t) = (1 - t)\mathbf{a} + t\mathbf{b}$$

Now $g = f \circ \mathbf{h}$ satisfies all the hypotheses of the single-variable mean-value theorem. So there exists t_0 in $(0,1)$ for which

$$g(1) - g(0) = g'(t_0)(1 - 0).$$

By the chain rule, $g'(t_0) = [\mathbf{D}f(\mathbf{h}(t_0))]D\mathbf{h}(t_0)$

Set $\mathbf{c} = \mathbf{h}(t_0)$ and this becomes

$$f(\mathbf{b}) - f(\mathbf{a}) = [\mathbf{D}f(\mathbf{c})](\vec{\mathbf{b} - \mathbf{a}}).$$

If points \mathbf{b} and \mathbf{a} differ only in their i th component, so that $\mathbf{b} = \mathbf{a} + \vec{\mathbf{e}}_i$, then

$$f(\mathbf{a} + \vec{\mathbf{e}}_i) - f(\mathbf{a}) = D_i f(\mathbf{a} + t_0 \vec{\mathbf{e}}_i), 0 < t_0 < 1.$$

13. (Proof 11.2) Using the mean value theorem, prove that if a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ has partial derivatives D_1f and D_2f that are continuous at \mathbf{a} , it is differentiable at \mathbf{a} and its derivative is the Jacobian matrix $[D_1f(\mathbf{a}) \ D_2f(\mathbf{a})]$.

14. Derivative of a function of a matrix (Example 1.7.17 in Hubbard):

A matrix is also a vector. When we square an $n \times n$ matrix A , the entries of $S(A) = A^2$ are functions of all the entries of A . If we change A by adding to it a matrix H of small length, we will make a change in the function value A^2 that is a linear function of H plus a small “remainder.”

We could in principle represent A by a column vector with n^2 components and the derivative of S by a very large matrix, but it is more efficient to leave H in matrix form and use matrix multiplication to find the effect of the derivative on a small increment matrix H . The derivative is still a linear function, but it is represented by matrix multiplication in a different way.

- (a) Using the definition of the derivative, show that the linear function that we want is $DS(H) = AH + HA$.
- (b) Confirm that DS is a linear function of H
- (c) Check that $DS(H)$ is a good approximation to $S(A+H) - S(A)$ for the following simple case, where the matrices A and H do not commute.

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, H = \begin{bmatrix} 0 & h \\ k & 0 \end{bmatrix}$$

15. Derivative of the matrix-inverse function

Define the function $T(A) = A^{-1}$. We expect that $T(A + H) - T(A)$ can be well approximated by an expression that is linear in H .

Proof strategy: use a geometric series

If we were dealing with numbers, we could write

$$\frac{1}{a+h} = \frac{1}{a} \frac{1}{1+\frac{h}{a}} = \frac{1}{a} \left(1 - \frac{h}{a} + \frac{h^2}{a^2} - \dots\right) =$$

This approach works with matrices, too, but we must be careful with the order of factors and remember that $(BC)^{-1} = C^{-1}B^{-1}$.

- Prove that $(A + H)^{-1} = (I + A^{-1}H)^{-1}A^{-1}$.
- Expand in a geometric series.
- Evaluate $T(A + H) - T(A)$ and identify the term that is linear in H .
Now we have our guess for the derivative.

- The “remainder” is $T(A + H) - T(A) + A^{-1}HA^{-1}$, and we have found that

$$T(A + H) = (A + H)^{-1} = A^{-1} - A^{-1}HA^{-1} + A^{-1}HA^{-1}HA^{-1} - A^{-1}HA^{-1}HA^{-1}HA^{-1} + \dots$$

Get a formula for this remainder that includes two factors of H . Then take its length. Use two strategies:

Length of product \leq product of lengths.

Length of sum \leq sum of lengths (generalized triangle inequality.)

- Prove that

$$\lim_{H \rightarrow 0} \frac{|Remainder|}{|H|} = 0.$$

16. Chain rule for functions of matrices

We have shown that the derivative of the squaring function $S(A) = A^2$ is $DS(H) = AH + HA$.

We also showed that for $T(A) = A^{-1}$, the derivative is $DT(H) = -A^{-1}HA^{-1}$.

Now the function $U(A) = A^{-2}$ can be expressed as the composition $U = S \circ T$.

Find the derivative $DU(H)$ by using the chain rule.

The chain rule says “the derivative of a composition is the composition of the derivatives,” even in a case like this where composition is not represented by matrix multiplication.

17. Newton's method

- (a) One variable: Function f is differentiable. You are trying to solve the equation $f(x) = 0$, and you have found a value a_0 , close to the desired x , for which $f(a_0)$ is small. Derive the formula $a_1 = a_0 - f(a_0)/f'(a_0)$ for an improved estimate.
- (b) Use Newton's method to find an approximate value for the cube root of 8.1.

- (c) n variables: U is an open subset of \mathbb{R}^n , and function $\vec{f}(\mathbf{x}) : U \rightarrow \mathbb{R}^n$ is differentiable. You are trying to solve the equation $\vec{f}(\mathbf{x}) = \vec{0}$, and you have found a value \mathbf{a}_0 , close to the desired \mathbf{x} , for which $\vec{f}(\mathbf{a}_0)$ is small. Derive the formula

$$\mathbf{a}_1 = \mathbf{a}_0 - [\mathbf{D}\vec{f}(\mathbf{a}_0)]^{-1}\vec{f}(\mathbf{a}_0).$$

for an improved estimate.

18. Newton's method – an example with two variables

We want an approximate solution to the equations

$$\log x + \log y = 3$$

$$x^2 - y = 1$$

$$\text{i.e. } f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \log x + \log y - 3 \\ x^2 - y - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Knowing that $\log 3 \approx 1.1$, show that $\mathbf{x}_0 = \begin{pmatrix} 3 \\ 9 \end{pmatrix}$ is an approximate solution to this equation, then use Newton's method to improve the approximation. Here is a check:

$$\log 2.81 + \log 6.87 = 2.98$$

$$2.81^2 - 6.87 = 1.02$$

19. Derivative of inverse function

Suppose that $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a continuously differentiable function. Choose a point \mathbf{x}_0 where the derivative $[\mathbf{D}\mathbf{f}(\mathbf{x}_0)]$ is an invertible matrix. Set $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$. Let \mathbf{g} be the differentiable local inverse function $\mathbf{g} = \mathbf{f}^{-1}$ such that $\mathbf{g}(\mathbf{y}_0) = \mathbf{x}_0$ and $\mathbf{f}(\mathbf{g}(\mathbf{y})) = \mathbf{y}$ if \mathbf{y} is close enough to \mathbf{y}_0 .

Prove that $[\mathbf{D}\mathbf{g}(\mathbf{y}_0)] = [\mathbf{D}\mathbf{f}(\mathbf{x}_0)]^{-1}$

20. An economic example of the inverse-function theorem:

Your model: Providing x in health benefits and y in educational benefits leads to happiness H and cost C according to the equation

$$\begin{pmatrix} H \\ C \end{pmatrix} = \mathbf{f} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + x^{0.5}y \\ x^{1.5} + y^{0.5} \end{pmatrix}.$$

Currently, $x = 4, y = 9, H = 22, C = 11$. Your budget is cut, and you are told to adjust x and y to reduce C to 10 and H to 19. Find an approximate solution by using the inverse-function theorem.

We cannot find formulas for the inverse function $\mathbf{g} \begin{pmatrix} H \\ C \end{pmatrix}$ that would solve the problem exactly, but we can calculate the derivative of \mathbf{g} .

(a) Check that $[\mathbf{D}\mathbf{f}] = \begin{bmatrix} 1 + \frac{y}{2\sqrt{x}} & \sqrt{x} \\ \frac{3}{2}\sqrt{x} & \frac{1}{2\sqrt{y}} \end{bmatrix} = \begin{bmatrix} \frac{13}{4} & 2 \\ 3 & \frac{1}{6} \end{bmatrix}$ is invertible.

(b) Use the derivative $[\mathbf{D}\mathbf{g}] = \begin{bmatrix} -0.03 & 0.36 \\ 0.55 & -0.6 \end{bmatrix}$ to approximate $\mathbf{g} \begin{pmatrix} 19 \\ 10 \end{pmatrix}$

21. Implicit functions – review of the linear case.

From linear algebra we are already familiar with the situation where a system of linear equations has more than one solution. This situation arises when there are fewer equations than unknowns.

We have n unknowns, $n - k$ equations, e.g for $n = 3, k = 1$

$$2x + 3y - 4z = 0$$

$$x + 2y - 3z = 0$$

Create an $(n - k) \times n$ matrix: $T = \begin{bmatrix} 2 & 3 & -4 \\ 1 & 2 & -3 \end{bmatrix}$

If the matrix T is not onto, its rows (the equations) are linearly dependent. Otherwise, when we row reduce, we will find $n - k = 2$ pivotal columns and $k = 1$ nonpivotal columns. We assign values arbitrarily to the “active” variables that correspond to the nonpivotal columns, and then the values of the “passive” variables that corresponds to the pivotal column are determined.

Suppose that we reorder the unknowns so that the “active” variables come last. Then, after we row reduce the matrix, the first $n - k$ columns will be pivotal. So the first $n - k$ columns will be linearly independent, and they form an invertible square matrix. The matrix is now of the form $T = [A|B]$, where A is invertible.

The solution vector is of the form $\vec{v} = \begin{bmatrix} \vec{x} \\ \vec{z} \end{bmatrix}$, where the passive variables \vec{x} come first, the active variables \vec{z} come second.

Substitute this form of \vec{v} into $T\vec{v} = \vec{0}$ and show that if we choose \vec{z} arbitrarily and set $\vec{x} = -A^{-1}B\vec{z}$, we have a solution.

Apply this technique to the given matrix and thereby get a solution where “passive” x and y are expressed in terms of “active” z .

Our system of equations determines \vec{x} “implicitly” in terms of \vec{z} .

22. The implicit function theorem

Let W be an open subset of \mathbb{R}^n , and let $\mathbf{F} : W \rightarrow \mathbb{R}^{n-k}$ be a C^1 mapping such that $\mathbf{F}(\mathbf{c}) = \mathbf{0}$. Assume that $[\mathbf{DF}(\mathbf{c})]$ is onto.

Prove that the n variables can be ordered so that the first $n - k$ columns of $[\mathbf{DF}(\mathbf{c})]$ are linearly independent, and that $[\mathbf{DF}(\mathbf{c})] = [A|B]$ where A is an invertible $(n - k) \times (n - k)$ matrix.

Set $\mathbf{c} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}$, where \mathbf{a} are the $n - k$ passive variables and \mathbf{b} are the k active variables.

Let \mathbf{g} be the “implicit function” from a neighborhood of \mathbf{b} to a neighborhood of \mathbf{a} such that $\mathbf{g}(\mathbf{b}) = \mathbf{a}$ and $\mathbf{F} \begin{pmatrix} \mathbf{g}(\mathbf{y}) \\ \mathbf{y} \end{pmatrix} = \mathbf{0}$.

Prove that $[\mathbf{Dg}(\mathbf{b})] = -A^{-1}B$.

23. Implicit functions – three variables, one equation

The nonlinear equation $F \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x^2 - 4z^2 - 4y^2 - 1 = 0$ implicitly deter-

mines x as a function of y and z .

If x is positive, $x = \sqrt{1 + 4y^2 + 4z^2}$; if x is negative, $x = -\sqrt{1 + 4y^2 + 4z^2}$.

Near the point $x = 3, y = 1, z = 1$, which is the appropriate explicit function

$$x = g \begin{pmatrix} y \\ z \end{pmatrix}?$$

There are three ways, in this simple example, to find the derivative of the implicitly defined function g . Two of them take advantage of the fact that there is only one constraint equation and only one active variable.

- (a) Solve algebraically to find a formula for the function g that expresses the passive variable x in terms of active variables y and z .

Then calculate $[Dg \begin{pmatrix} 1 \\ 1 \end{pmatrix}]$ for the appropriate function.

- (b) Use the implicit function theorem to calculate the partial derivatives of g from the partial derivatives of F :

Calculate the Jacobian matrix $[DF]$ at $x = 3, y = 1, z = 1$.

Split off a square matrix A on the left, so that $[DF] = [A|B]$, and confirm that $[Dg \begin{pmatrix} 1 \\ 1 \end{pmatrix}] = -A^{-1}B$.

- (c) Since there is only one passive variable x , just replace x by $g \begin{pmatrix} y \\ z \end{pmatrix}$ to

$$\text{get } h \begin{pmatrix} y \\ z \end{pmatrix} = g \begin{pmatrix} y \\ z \end{pmatrix}^2 - 4z^2 - 4y^2 - 1 = 0.$$

Take the partial derivative $D_1 h \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, using the chain rule, and solve

for $D_1 g \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. In this context the notation $\frac{\partial g}{\partial y}$ is standard.

24. Finding the derivative of an implicitly defined function

The plane $x + 2y - 3z + 4 = 0$ and the cone $x^2 + y^2 - z^2 = 0$ intersect in a curve that includes the point $\mathbf{c} = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$. Near that point this curve is the graph of a function $\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{g}(z)$.

Use the implicit function theorem to determine $\mathbf{g}'(5)$, then find the approximate coordinates of a point on the curve with $z = 5.1$.

Check: $2.9 + 2(4.2) - 3(5.1) + 4 = 0$; $2.9^2 + 4.2^2 - 5.1^4 = 0.04$.

25. Curves, Surfaces, Graphs, and Manifolds

Manifolds are a generalization of smooth curves and surfaces to an arbitrary number of dimensions.

The simplest sort of manifold is a flat one. Consider an affine subset $X \subset \mathbb{R}^n$. This subset could be a subspace, or it could be a subspace plus a constant, like a line or plane that does not include the origin.

There are three equivalent ways to describe such a manifold. The first will generalize to the definition of a manifold, but the second and third are more common and useful. Try them out on the line of slope 2 that passes through the point $x = 0, y = -2$.

(a) As the **graph** of a function. There are two ways to do this!

(b) As the **locus** defined by an equation $F\begin{pmatrix} x \\ y \end{pmatrix} = 0$.

(c) By a **parametrization** in term of a point and a direction vector, using a parameter t .

Graphs and manifolds in \mathbb{R}^3 and \mathbb{R}^n

Example 1: Consider the graph of $f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - 4$

This graph is the plane $2x + 3y - z = 4$. It is a subset of the space \mathbb{R}^3 whose dimension, $n = 3$, is the sum of the dimension of the domain of f , $k = 2$, and the dimension of the codomain of f , $n - k = 1$.

Describe this manifold as the locus of a function $F : \mathbb{R}^3 \rightarrow \mathbb{R}$.

Invent a parametric description of this manifold.

Example 2: Consider the graph of $\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{f}(z) = \begin{bmatrix} 3 + 2z \\ 3 - z \end{bmatrix}$

This graph is a line.

Invent a parametric description of this manifold.

Invent a description of this manifold as a locus.

Again, the manifold is a subset of the space \mathbb{R}^3 whose dimension, $n = 3$, is the sum of the dimension of the domain of f , $k = 1$, and the dimension of the codomain of f , $n - k = 2$.

Make the function $\mathbf{f} : \mathbb{R}^k \rightarrow \mathbb{R}^{n-k}$ arbitrary and we have a graph Γ :

the set of pairs (\mathbf{x}, \mathbf{y}) such that $\mathbf{y} = \mathbf{f}(\mathbf{x})$. This is a subset of \mathbb{R}^n .

Require that the function be C^1 (continuously differentiable) and we almost have a manifold. But it is asking too much for a single function always to do the job.

26. Definition, and examples in \mathbb{R}^3

A subset $M \subset \mathbb{R}^n$ is a smooth manifold if locally it is the graph of a C^1 function (the partial derivatives are continuous). “Locally” means that for any point $\mathbf{x} \in M$ we can find a neighborhood U of \mathbf{x} such that within $M \cap U$, there is a C^1 function that expresses $n - k$ variables in terms of the other k . The number k is the dimension of the manifold. Finding such a function is precisely the job of the implicit function theorem.

- $k = 3$. Any open subset $M \subset \mathbb{R}^3$ is a smooth 3-dimensional manifold. In this case $k = 3$, and the manifold is the graph of a function $f : \mathbb{R}^3 \rightarrow \{\vec{0}\}$, whose codomain is the trivial vector space $\{\vec{0}\}$ that contains just a single point. Such a function is necessarily constant, and its derivative is zero.

Why does M have to be open?

- $k = 2$. Consider the graph of $f \begin{pmatrix} x \\ y \end{pmatrix} = x^2 + y^2$.

What is the name of this surface?

- $k = 1$. Consider the graph of the function $\vec{f}(z) = \begin{pmatrix} \cos 2\pi z \\ \sin 2\pi z \end{pmatrix}$. What is the name of this curve?

- $k = 0$. In this case the manifold consists of one or more isolated points. Near any of these points \mathbf{x}_0 , it is the graph of a function $\vec{f} : \{\vec{0}\} \rightarrow \mathbb{R}^3$ whose domain is the trivial vector space and whose image is the point $\mathbf{x}_0 \in \mathbb{R}^3$.

This function is differentiable. Its derivative is the zero function. Whatever $\epsilon > 0$ you choose, $\vec{f}(\vec{0} + \vec{h}) - \vec{f}(\vec{0}) < \epsilon$ for all \vec{h} , since the only possible \vec{h} is $\vec{0}$.

27. Manifolds defined by geometry

For the unit circle, we need four functions to implement the preceding definition.

Show how to break up the circle into four pieces, each of which is a function graph.

What single equation can characterize the unit circle as a locus?

For the unit sphere, we need six functions in place of the single equation $x^2 + y^2 + z^2 = 1$.

How would you slice up an apple to illustrate this?

28. Other examples of manifolds

- An “indifference” surface. Let x = days of vacation per year, y = fraction of health insurance paid by the company, z = number of shares of stock options available for purchase annually. The manifold M consists of all points $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ that employees regard as neither better nor worse than $\begin{pmatrix} 15 \\ 0.5 \\ 100 \end{pmatrix}$.
- We are given as \mathbf{F} three equations involving five variables, so that $n = 5, k = 2$. In this case $M \cap U$ might be part of a two-dimensional manifold, a surface in \mathbb{R}^5 . In economics this is easy to picture: choose five variables that specify how the Federal budget is divided up among Cabinet departments, and invent three equations that impose constraints specified by the Pentagon, the White House, and the Congressional budget office. Three of the budget variables become “passive”: they are functions of the other two “active” variables.
- The set of equilibrium states of one mole of helium gas confined to a cylinder fitted with a movable piston. There are many pairs of variables that specify the state (e.g. pressure and volume, or temperature and entropy). This is a 2-dimensional manifold, but it is not an obvious subset of some \mathbb{R}^n .
- The set of positions for a pair of particles joined by a rigid rod of length l . Specify x_1, y_1 , and x_2 , with $|x_2 - x_1| \leq l$, and you can calculate y_2 , but you need to know whether $y_2 - y_1$ is positive or negative. Make a sketch to illustrate the problem.
- The configuration of a set of four particles joined into a quadrilateral by four rigid rods. This instructive example is discussed in detail on pp. 295-296 of Hubbard, and it provides the basis for some interesting homework problems. Sketch an example to show how the location of one joint depends on the others, but only locally.

29. Identifying a smooth manifold by the implicit function theorem

This is Hubbard Theorem 3.1.10.

We have a subset $M \subset \mathbb{R}^n$ defined as a “locus” and would like to show that M is a smooth k -dimensional manifold.

Suppose there is an open subset $U \subset \mathbb{R}^n$ and a C^1 function $\mathbf{F} : U \rightarrow \mathbb{R}^{n-k}$ such that the “locus,” the set of solutions of the equation $\mathbf{F}(\mathbf{z}) = 0$, is $M \cap U$.

If $[\mathbf{DF}(\mathbf{z})]$ is onto (surjective) for every $\mathbf{z} \in M \cap U$, then $M \cap U$ is a smooth k -dimensional manifold embedded in \mathbb{R}^n .

Make a sketch to illustrate $M \cap U$.

To say that M itself is a manifold, we have to find such a U for every point \mathbf{z} in the manifold, perhaps with different functions at different points.

Suppose that M consists of a circle of radius 1 whose center is the origin and a circle of radius 1 centered at $x = 3, y = 0$. What pair of choices for F and U will define the entire manifold?

Proof: the implicit function theorem says precisely this. The statement that $[\mathbf{DF}(\mathbf{z})]$ is onto guarantees the differentiability of the implicitly defined function.

30. Testing that a locus is a manifold (Example 3.1.11)

$F \begin{pmatrix} x \\ y \end{pmatrix} = x^8 + 2x^3 + y + y^5 - c = 0$. The locus is a smooth manifold, which looks surprisingly normal when plotted for various values of c .

Prove that $[DF]$ is onto for any value of c .

If you want to represent the manifold as a graph, which should be the active variable?

Is there a function $y = f(x)$ that describes the manifold? Can you find a formula for it?

31. Testing that a locus is a manifold (Example 3.1.12)

$F\begin{pmatrix} x \\ y \end{pmatrix} = x^4 + y^4 + x^2 - y^2 = c$. Construct $[\mathbf{D}F]$, and find the three points where it fails to be onto because both partial derivatives vanish. For each point, choose c so that the point is on the locus.

Here is a plot, generated in R, for various positive values of c .

Here is a plot, generated in R, for various negative values of c .

3 Seminar Topics

Your section instructor will either have emailed a list of topics to prepare or will have posted a sign-up list of appointments on the Calendar tab of Canvas. Either way, there will be one of the following topics that you should be prepared to present.

Practice your presentation so that it takes about 8 minutes. The text of the presentation will be projected onto a screen so that you need not recopy it. To save time, avoid writing long sentences on the chalkboard. You may use notes, but be discreet about it.

1. (Proof 11.1) Let $U \subset \mathbb{R}^n$ be an open set, and let f and g be functions from U to \mathbb{R} . Prove that if f and g are differentiable at \mathbf{a} then so is fg , and that

$$[\mathbf{D}(fg)(\mathbf{a})] = f(\mathbf{a})[\mathbf{D}g(\mathbf{a})] + g(\mathbf{a})[\mathbf{D}f(\mathbf{a})].$$

2. Draw a diagram like the one on page 13 to illustrate how the chain rule works for functions from \mathbb{R}^2 to \mathbb{R}^2 .

$U \subset \mathbb{R}^2$ (on the left), $V \subset \mathbb{R}^2$ (in the middle), and $W \subset \mathbb{R}^2$ (on the right) are open sets, \mathbf{a} is a point in U , and $\vec{\mathbf{h}}$ is a small increment vector attached to \mathbf{a} .

$\mathbf{g} : U \rightarrow V$ is differentiable at \mathbf{a} , and its derivative $[\mathbf{D}\mathbf{g}(\mathbf{a})]$ is a 2×2 Jacobian matrix. You can use this Jacobian to find a good approximation to the increment vector that takes you in V from $\mathbf{g}(\mathbf{a})$ to $\mathbf{g}(\mathbf{a} + \vec{\mathbf{h}})$

$\mathbf{f} : V \rightarrow W$ is differentiable at $\mathbf{g}(\mathbf{a})$, and its derivative there, $[\mathbf{D}\mathbf{f}(\mathbf{g}(\mathbf{a}))]$, is a 2×2 Jacobian matrix. You can apply this Jacobian to the increment vector in V and thereby find a good approximation to the increment vector in W , $(\mathbf{f} \circ \mathbf{g})(\mathbf{a} + \vec{\mathbf{h}}) - (\mathbf{f} \circ \mathbf{g})(\mathbf{a})$,

You could alternatively skip over V and apply the Jacobian $[\mathbf{D}(\mathbf{f} \circ \mathbf{g})(\mathbf{a})]$ to $\vec{\mathbf{h}}$ to get a good approximation to the increment in W .

Show that equating these two approximations leads to the chain rule. What makes the proof messy is showing that the difference between the approximations goes rapidly to zero when $|\vec{\mathbf{h}}|$ is small – do not try!

3. (Proof 11.2) Using the mean value theorem, prove that if a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ has partial derivatives D_1f and D_2f that are continuous at \mathbf{a} , it is differentiable at \mathbf{a} and its derivative is the Jacobian matrix $[D_1f(\mathbf{a}) \ D_2f(\mathbf{a})]$.

4. (The inverse function theorem)

Review: suppose that $f : \mathbb{R} \rightarrow \mathbb{R}$ is differentiable at x_0 , $g : \mathbb{R} \rightarrow \mathbb{R}$ is its differentiable inverse function, and $y_0 = f(x_0)$. Apply the chain rule to show that $g'(y_0) = 1/f'(x_0)$.

Now do the same thing in two dimensions. Suppose that $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a continuously differentiable function. Choose a point \mathbf{x}_0 where the derivative $[\mathbf{Df}(\mathbf{x}_0)]$ is an invertible matrix. Set $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$. Let \mathbf{g} be the differentiable local inverse function $\mathbf{g} = \mathbf{f}^{-1}$ such that

$\mathbf{g}(\mathbf{y}_0) = \mathbf{x}_0$ and $\mathbf{f}(\mathbf{g}(\mathbf{y})) = \mathbf{y}$ if \mathbf{y} is close enough to \mathbf{y}_0 .

Again use the chain rule to prove that $[\mathbf{Dg}(\mathbf{y}_0)] = [\mathbf{Df}(\mathbf{x}_0)]^{-1}$

5. (Proof 11.3) The implicit function theorem

Let W be an open subset of \mathbb{R}^n , and let $\mathbf{F} : W \rightarrow \mathbb{R}^{n-k}$ be a C^1 mapping such that $\mathbf{F}(\mathbf{c}) = \mathbf{0}$. Assume that $[\mathbf{DF}(\mathbf{c})]$ is onto.

Prove that the n variables can be ordered so that the first $n - k$ columns of $[\mathbf{DF}(\mathbf{c})]$ are linearly independent, and that $[\mathbf{DF}(\mathbf{c})] = [A|B]$ where A is an invertible $(n - k) \times (n - k)$ matrix.

Set $\mathbf{c} = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}$, where \mathbf{a} are the $n - k$ passive variables and \mathbf{b} are the k active variables.

Let \mathbf{g} be the “implicit function” from a neighborhood of \mathbf{b} to a neighborhood of \mathbf{a} such that $\mathbf{g}(\mathbf{b}) = \mathbf{a}$ and $\mathbf{F} \begin{pmatrix} \mathbf{g}(\mathbf{y}) \\ \mathbf{y} \end{pmatrix} = \mathbf{0}$.

Prove that $[\mathbf{Dg}(\mathbf{b})] = -A^{-1}B$.

6. (Extra topic) Here is a simple example of a non-differentiable function:

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \frac{x^2 y}{x^2 + y^2}; f \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0.$$

Prove that this function is not differentiable at $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ by using the definition of the directional derivative,

$$\nabla_{\vec{v}} f(\mathbf{a}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h\vec{v}) - f(\mathbf{a})}{h}$$

to calculate the directional derivative at $\mathbf{a} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ for the three unit vectors

$$\vec{\mathbf{v}}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \vec{\mathbf{v}}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \vec{\mathbf{v}}_3 = \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{pmatrix}$$

and concluding that the directional derivative is not a linear function of $\vec{\mathbf{v}}$.

4 Workshop Problems

1. Chain rule

(a) Chain rule for matrix functions

On page 23 of the lecture outline, we obtained the differentiation formula for $U(A) = A^{-2}$ by writing $U = S \circ T$ with $S(A) = A^2, T(A) = A^{-1}$. Prove the same formula from the chain rule in a different way, by writing $U = T \circ S$. You may reuse the formulas for the derivatives of S and T :

If $S(A) = A^2$ then $[DS(A)](H) = AH + HA$.

If $T(A) = A^{-1}$ then $[DT(A)](H) = -A^{-1}HA^{-1}$.

(b) Chain rule with 2×2 matrices

Start with a pair of polar coordinates $\begin{pmatrix} r \\ \theta \end{pmatrix}$.

Function \mathbf{g} converts them to Cartesian $\begin{pmatrix} x \\ y \end{pmatrix}$.

Function \mathbf{f} then converts $\begin{pmatrix} x \\ y \end{pmatrix}$ to $\begin{pmatrix} 2xy \\ x^2 - y^2 \end{pmatrix}$.

Confirm that $[\mathbf{D}(\mathbf{f} \circ \mathbf{g})\left(\begin{pmatrix} r \\ \theta \end{pmatrix}\right)] = [\mathbf{D}\mathbf{f}(\mathbf{g}\left(\begin{pmatrix} r \\ \theta \end{pmatrix}\right))] \circ [\mathbf{D}\mathbf{g}\left(\begin{pmatrix} r \\ \theta \end{pmatrix}\right)]$

2. Issues of differentiability

- (a) Suppose that A is a matrix and S is the cubing function given by the formula $S(A) = A^3$. Prove that S is differentiable and that its derivative is the linear function of the matrix H given by the formula $[DS(A)](H) = A^2H + AHA + HA^2$.

The proof consists in showing that the length of the “remainder” goes to zero faster than the length of the matrix H .

- (b) A continuous but non-differentiable function

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \frac{x^2y}{x^2 + y^2}, f\begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0.$$

- i. As in seminar topic 6, show that both partial derivatives vanish at the origin, so that the Jacobian matrix at the origin is the zero matrix $[0 \ 0]$, but that the directional derivative along $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is not zero. How does this calculation show that the function is not differentiable at the origin?
- ii. For all points except the origin, the partial derivatives are given by the formulas

$$D_1f\begin{pmatrix} x \\ y \end{pmatrix} = \frac{2xy^3}{(x^2 + y^2)^2}, D_2f\begin{pmatrix} x \\ y \end{pmatrix} = \frac{x^4 - x^2y^2}{(x^2 + y^2)^2}$$

Construct a “bad sequence” of points approaching the origin to show that D_1f is discontinuous at the origin.

3. Inverse functions and Newton's method

- (a) An approximate solution to the equations

$$x^3 + y^2 - xy = 1.08$$

$$x^2y + y^2 = 2.04$$

is $x_0 = 1$, $y_0 = 1$.

Use one step of Newton's method to improve this approximation.

- (b) You are in charge of building the parking lots for a new airport. You have ordered from amazon.com enough asphalt to pave 1 square kilometer, plus 5.6 kilometers of chain-link fencing. Your plan is to build two square, fenced lots. The long-term lot is a square of side $x=0.8$ kilometers; the short-term lot is a square of side $y=0.6$ kilometers. The amount of asphalt A and the amount C of chain-link fencing required are then specified by the function

$$\begin{pmatrix} A \\ C \end{pmatrix} = F \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 + y^2 \\ 4x + 4y \end{pmatrix},$$

Amazon decides to be generous. They deliver enough asphalt to pave 1.0032 square kilometers and 5.616 kilometers of fence.

- i. Use the inverse-function theorem to find approximate new values for x and y that use what was shipped to you.
- ii. Find a case where $A = 1$ but the value of C is such that this approach will fail because $[DF]$ is not onto. (This case corresponds to the maximum amount of fencing.)

4. Problems to be solved using R (if your group has experience with R, solve one of these as your third problem.)

Both problems can be done by modifying R script 3.3B. Since Newton's method is the only generally applicable technique for solving nonlinear equations, it is really useful to know how to do it on a computer.

(a) Saving Delos

The ancient citizens of Delos, threatened with a plague, consulted the oracle of Delphi, who told them to construct a new cubical altar to Apollo whose volume was double the size of the original cubical altar. (For details, look up "Doubling the cube" on Wikipedia.)

If the side of the original altar was 1, the side of the new altar had to be the real solution to $f(x) = x^3 - 2 = 0$.

Numerous solutions to this problem have been invented. One uses a "marked ruler" or "neusis"; another uses origami.

Your job is to use multiple iterations of Newton's method to find an approximate solution for which $x^3 - 2$ is less than 10^{-8} in magnitude.

(b) An approximate solution to the system of nonlinear equations

$$x + y^2 + z^3 = 9$$

$$xy + xz + yz = 12$$

$$xyz = 7$$

is $x = 3, y = 2, z = 1$.

Use two iterations of Newton's method to find a good approximate solution to these equations.

5 Homework(due Dec. 4, but do problems 1-4 by Nov. 27)

- (a) Hubbard, Exercise 1.7.21 (derivative of the determinant function). This is really easy if you work directly from the definition of the derivative.
- (b) Generalize this result to the 3×3 case. Hint: consider a matrix whose columns are $\vec{e}_1 + h\vec{a}_1$, $\vec{e}_2 + h\vec{a}_2$, $\vec{e}_3 + h\vec{a}_3$, and use the definition of the determinant as a triple product.
- Chain rule: an example with 2×2 matrices
A similar example with a 3×3 matrix is on page 151 of Hubbard.

The function

$\mathbf{f} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(x+y) \\ \sqrt{xy} \end{pmatrix}$ was invented by Gauss about 200 years ago to deal with integrals of the form

$$\int_{-\infty}^{\infty} \frac{dt}{\sqrt{(t^2 + x^2)(t^2 + y^2)}}.$$

It was revived in the late 20th century as the basis of the AGM (arithmetic-geometric mean) method for calculating π . You can get 1 million digits with a dozen or so iterations.

The function is meant to be composed with itself; so it will be appropriate to compute the derivative of $\mathbf{f} \circ \mathbf{f}$ by the chain rule.

- (a) \mathbf{f} is differentiable whenever x and y are positive; so its derivative is given by its Jacobian matrix. Calculate this matrix.

We choose to evaluate the derivative of $\mathbf{f} \circ \mathbf{f}$ at the point $\begin{pmatrix} 8 \\ 2 \end{pmatrix}$.

Conveniently, $\mathbf{f} \begin{pmatrix} 8 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$. The chain rule says that

$$[\mathbf{D}(\mathbf{f} \circ \mathbf{f})] \begin{pmatrix} 8 \\ 2 \end{pmatrix} = [\mathbf{D}\mathbf{f} \begin{pmatrix} 5 \\ 4 \end{pmatrix}][\mathbf{D}\mathbf{f} \begin{pmatrix} 8 \\ 2 \end{pmatrix}].$$

Evaluate the two numerical Jacobian matrices. Because the derivative of \mathbf{f} is evaluated at two different points, they will not be the same.

- (b) Write the formula for $\mathbf{f} \circ \mathbf{f}$, compute and evaluate the lower left-hand entry in its Jacobian matrix, and check that it agrees with the value given by the chain rule.

3. Hubbard, Exercise 1.8.6, part (b) only. In the case where \mathbf{f} and \mathbf{g} are functions of time t , this formula finds frequent use in physics. You can either do the proof as suggested in part (a) or model your proof on the one for the dot product on page 143.
4. (similar to the second example on page 14)
Hubbard, Exercise 1.8.9. The equation that you prove can be called a “first-order partial differential equation.”
5. (similar to workshop problem 1a)
We know the derivatives of the matrix-squaring function S and the matrix-inversion function T :
If $S(A) = A^2$ then $[DS(A)](H) = AH + HA$.
If $T(A) = A^{-1}$ then $[DT(A)](H) = -A^{-1}HA^{-1}$.
 - (a) Use the chain rule to find a formula for the derivative of the function $U(a) = A^4$.
 - (b) Use the chain rule to find a formula for the derivative of the function $W(a) = A^{-4}$.
6. (a) Hubbard, problem 2.10.2. Make a sketch to show how this mapping defines an alternative coordinate system for the plane, in which a point is defined by the intersection of two hyperbolas.
 - (b) The point $x = 3, y = 2$ is specified in this new coordinate system by the coordinates $u = 6, v = 5$. Use the derivative of the inverse function to find approximate values of x and y for a nearby point where $u = 6.5, v = 4.5$. (This is essentially one iteration of Newton’s method.)
 - (c) Find h such that the point $u = 6 + h, v = 5.1$ has nearly the same x -coordinate as $u = 6, v = 5$.
 - (d) Find k such that the point $x = 3 + k, y = 2.1$ has nearly the same u -coordinate as $x = 3, y = 2$.
 - (e) For this mapping, you can actually find a formula for the inverse function that works in the region of the plane where x, y, u , and v are all positive. Find the rather messy formulas for x and y as functions of u and v , and use them to answer the earlier questions. Once you calculate the Jacobian matrix and plug in appropriate numerical values, you will be back on familiar ground.

7. (similar to workshop problem 2b)

As a summer intern, you are given the job of reconciling the Democratic and Republican proposals for tax reform. Both parties agree on the following model:

- x is the change in the tax rate for the middle class.
- y is the change in the tax rate for the well-off.
- The net impact on revenue is given by the function

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \frac{x(x^2 - y^2)}{x^2 + y^2}, f \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0.$$

The Republican proposal is $y = -x$, while the Democratic proposal is $y = x$.

- (a) Show that f is continuous at the origin.
- (b) Show that both proposals are revenue neutral by calculating two appropriate directional derivatives. You will have to use the definition of the directional derivative, not the Jacobian matrix.
- (c) At the request of the White House, you investigate a 50-50 mix of the two proposals, the compromise case where $y = 0$, and you discover that it is not revenue neutral! Confirm this surprising conclusion by showing that the directional derivatives at the origin cannot be given by a linear function; i.e. that f is not differentiable.
- (d) Your final task is to explain the issue in terms that legislators can understand: the function is not differentiable because its partial derivatives are not continuous. Demonstrate that one of the partial derivatives of f is discontinuous at the origin. (D_2f is less messy.)

8. The CEO of a chain of retail stores will get a big bonus if she hits her volume and profit targets for December exactly. Her microeconomics consultant, fresh out of Harvard, tells her that both her target figures are functions of two variables, investment x in Internet advertising and investment y in television advertising. The former attracts savvier customers and so tends to contribute to volume more than to profit.

The function that determines volume V and profit P is

$$\begin{pmatrix} V \\ P \end{pmatrix} = \begin{pmatrix} x^{\frac{3}{4}}y^{\frac{1}{3}} + x \\ x^{\frac{1}{4}}y^{\frac{2}{3}} + y \end{pmatrix}.$$

With $x = 16, y = 8, V = 32, P = 16$, our CEO figures she is set for a big bonus. Suddenly, the board of directors, feeling that Wall Street is looking as much for profit as for volume this year, changes her targets to $V = 24, P = 24$. She needs to modify x and y to meet these new targets.

Near $V = 32, P = 16$, there is an inverse function such that

$\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{g} \begin{pmatrix} V \\ P \end{pmatrix}$. Find its derivative $[\mathbf{D}\mathbf{g}]$, and use the derivative to find values of x and y that are an approximate solution to the problem. Because the increments to V and P are large, you should not expect the approximate solution to be very good, but it will be better than doing nothing.

9. (Hubbard, exercise 3.12)

Let $X \subset \mathbb{R}^3$ be the set of midpoints of segments joining a point of the curve C_1 of equation $y = x^2, z = 0$ to a point of the curve C_2 of equation $z = y^2, x = 0$.

- (a) Parametrize C_1 and C_2 .
- (b) Parametrize X .
- (c) Find an equation for X (i.e. describe X as a locus)
- (d) Show that X is a smooth surface.

10. Manifold M is known by the equation $F \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x^2 + y^4 - 2z^2 - 2 = 0$ near

the point $\mathbf{c} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$.

- (a) Locally, near \mathbf{c} , M is the graph of a function $x = g \begin{pmatrix} y \\ z \end{pmatrix}$. Determine $[\mathbf{D}g(\mathbf{c})]$ by using the implicit function theorem.
- (b) Use $[\mathbf{D}g(\mathbf{c})]$ to find the approximate value of x for a point of M near \mathbf{c} for which $y = 1.1, z = 1.8$.
- (c) Check your answers by finding an explicit formula for g and taking its derivative.

Optional extra problems to be solved using R

11. In problem 8, use multiple iterations of Newton's method in R to find accurate values of x and y that meet the revised targets. Feel free to modify Script 3.3C.

12. (Related to group problem 4a)

The quintic equation $x(x^2 - 1)(x^2 - 4) = 0$ clearly has five real roots that are all integers. So does the equation $x(x^2 - 1)(x^2 - 4) - 1 = 0$, but you have to find them numerically. Get all five roots using Newton's method, carrying out enough iterations to get an error of less than .001. Use R to do Newton's method and to check your answers. If you have R plot a graph, it will be easy to find an initial guess for each of the five roots.