Mathematical Analysis Outline

William G. Faris

January 8, 2007
Chapter 1

Metric spaces and continuous maps

1.1 Metric spaces

A metric space is a set $X$ together with a real distance function $(x, x') \mapsto d(x, x') \geq 0$. This distance function must satisfy the following axioms:

1. $d(x, x) = 0$,
2. $d(x, z) \leq d(x, y) + d(y, z)$,
3. $d(x, y) = d(y, x)$,
4. $d(x, y) = 0$ implies $x = y$.

One way to get a metric space is from a normed vector space. This is a vector space together with a real norm function $\|x\| \geq 0$. This norm function must satisfy the following axioms:

1. $\|cu\| = c\|u\|$,
2. $\|u + v\| \leq \|u\| + \|v\|$,
3. $\|u\| = 0$ implies $\|u\| = 0$.

Every normed vector space is a metric space, if we define $d(u, v) = \|u - v\|$. Here are some basic examples of normed vector spaces.

1. $\mathbb{R}^n$ with the $\ell_1^n$ norm $\|x\|_1 = \sum_{i=1}^n |x_i|$.
2. $\mathbb{R}^n$ with the $\ell_2^n$ norm $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$.
3. $\mathbb{R}^n$ with the $\ell_\infty^n$ norm $\|x\|_\infty = \sup_{1 \leq i \leq n} |x_i|$.
4. The space $\ell^1$ of infinite real sequences such that the $\ell^1$ norm $\|x\|_1 = \sum_{i=1}^{\infty} |x_i| < +\infty$.
5. The space $\ell^2$ of infinite real sequences such that the $\ell^2$ norm $\|x\|_2 = \sqrt{\sum_{i=1}^{\infty} x_i^2} < +\infty$.
6. The space $\ell^\infty$ of infinite real sequences such that the $\ell^\infty$ norm $\|x\|_\infty = \sup_{1 \leq i < +\infty} |x_i| < +\infty$.
7. The space $C(K)$ of continuous real functions on a compact metric space $K$ with the uniform norm $\|f\| = \max_{t \in K} |f(t)|$.

The relations between the $\ell^p_n$ norms are $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq n\|x\|_\infty$. Furthermore, $\|x\|_1 \leq \sqrt{n}\|x\|_2$ and $\|x\|_2 \leq \sqrt{n}\|x\|_\infty$.

The relations between the $\ell^p$ norms are $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$. So $\ell^1 \subset \ell^2 \subset \ell^\infty$.

Every subset of a metric space is also a metric space, if we restrict the distance function to points in the subset.

If $x$ is a point in a metric space $M$, then $B_r(x) = \{ y \in M \mid d(x,y) < r \}$.

If $U \subset M$, then $U$ is open if and only if for every $x \in U$ there is a $r > 0$ such that $B_r(x) \subset U$.

If $F \subset M$, then $F$ is closed if and only if every point $x$ such that for every $r > 0$ there is a point of $F$ in $B_r(x)$ is itself in $F$.

Open and closed subsets are complements of each other.

The union of a collection of open sets is open. The intersection of a collection of closed sets is closed.

The finite intersection of a collection of open sets is open. The finite union of a collection of closed sets is closed.

If $A$ is a subset of $M$, then $U \subset A$ is (relatively) open if and only if there is an open subset $V$ of $M$ such that $V \cap M = U$. If $B$ is a subset of $M$, then $F \subset A$ is (relatively) closed if and only if there is a closed subset $G$ of $M$ such that $G \cap M = F$.

1.2 Continuity

A sequence is a function $n \to s_n$ from $\mathbb{N}$ to $M$. A sequence converges to a limit $x$, or $s_n \to x$, provided that $\forall \epsilon > 0 \exists N \forall n \geq N \ d(s_n, x) < \epsilon$.

A subset $F \subset M$ is closed if and only if every sequence $n \to s_n$ of points in $F$ that converges has its limit also in $F$.

If $f : X \to Y$ is a function from one metric space to another, then $f$ is continuous if for every $x$ in $X$ we have $\forall \epsilon > 0 \exists \delta > 0 \forall x' (d(x', x) < \delta \Rightarrow d(f(x'), f(x)) < \epsilon)$.

A function $f : X \to Y$ is continuous if and only whenever $s_n \to x$ in $X$, then $f(s_n) \to f(x)$ in $Y$.

A function $f : X \to Y$ is continuous if and only whenever $V$ is an open subset of $Y$, then $f^{-1}(V)$ is an open subset of $X$. 
1.3. CONNECTED SPACES

If \( f : X \to Y \) is a function from one metric space to another, then \( f \) is uniformly continuous if \( \forall \varepsilon > 0 \exists \delta > 0 \forall x', x (d(x', x) < \delta \Rightarrow d(f(x'), f(x)) < \varepsilon) \).

A function \( f : X \to Y \) is uniformly continuous if and only whenever \( d(s_n, t_n) \to 0 \), then \( d(f(s_n), f(t_n)) \to 0 \).

1.3 Connected spaces

A metric space \( A \) is connected if it is not the disjoint union of two non-empty open subsets.

If \( A \) is connected, and \( f : A \to B \) is continuous, then \( f(A) \) is connected.

A subset \( C \) of \( \mathbb{R} \) is connected if and only if it is an interval, that is, \( x, z \) in \( C \) and \( x < y < z \) implies \( y \) is in \( C \).

Proof: Suppose \( A \) not an interval, then there are \( x, z \) in \( C \) and \( x < y < z \) with \( y \) not in \( C \). The part of \( A \) below \( y \) and the part of \( A \) above \( y \) disconnect \( A \). So \( A \) is not connected.

Suppose, on the other hand, that \( A \) is not connected. Suppose that \( U \) and \( V \) are disjoint non-empty open subsets of \( A \). Suppose that \( x \) in in \( U \) and \( z \) is in \( V \) and \( x < z \). Consider the subset of \( A \) consisting of the points \( p \) with \( p \) in \( U \) and \( p < z \). This is non-empty and bounded above, so it has a supremum \( y \). Since \( y \) is an upper bound for \( U \), it cannot be in \( U \), since \( U \) is open. On the other hand, since \( y \) is a least upper bound, it cannot be in \( V \). Furthermore, it is clear that \( x < y < z \). So \( A \) is not an interval.

1.4 Complete metric spaces

A Cauchy sequence is a sequence such that \( \forall \varepsilon > 0 \exists N \forall m \geq N \forall n \geq N d(s_m, s_n) < \varepsilon \).

A metric space \( M \) is complete if every Cauchy sequence of points in \( M \) converges to a limit in \( M \).

A normed vector space that is also a complete metric space is called a Banach space. Each of the metric spaces in the catalog of normed vector spaces given above is a Banach space.

If \( M \) is a complete metric space, and if \( A \subset M \), then \( A \) is a complete metric space if and only if \( A \) is a closed subset of \( M \).

Contraction mapping theorem. Let \( M \) be a complete metric space. Let \( c < 1 \). Suppose that \( g : M \to M \) satisfies the Lipschitz condition \( d(g(x), g(y)) \leq cd(x, y) \). Then there is a unique solution of \( g(x) = x \). If \( x_0 \) is in \( M \) and if \( x_{n+1} = g(x_n) \), then the sequence \( x_n \) converges to this solution.

Solving equations by iteration. Say that \( f \) is a real continuous function defined on a closed interval \([x_0 - r, x_0 + r]\). Suppose that there is a number \( a \neq 0 \) such that \( g(x) = a^{-1}(ax - f(x)) \) is a contraction with Lipschitz constant \( 1/2 \). Suppose also that \( f(x_0) = y_0 \), and \( y \) satisfies \( |a^{-1}(y - y_0)| \leq r/2 \). Then \( f(x) = y \) has a unique solution in the interval.
CHAPTER 1. METRIC SPACES AND CONTINUOUS MAPS

Proof: Let
\[ g(x) = x - a^{-1}(f(x) - y). \]  \hspace{1cm} (1.1)
Then \( g \) is also Lipschitz with constant \( 1/2 \). Furthermore \( g(x) - x_0 = g(x) - g(x_0) + a^{-1}(y_0 - y) \). This has magnitude at most \( (1/2)|x - x_0| + r/2 \leq r \). So \( |x - x_0| \leq r \) implies \( |g(x) - x_0| \leq r \). Thus we have a strict contraction on this closed interval to which the contraction mapping theorem applies.

Inverse function theorem. Suppose that \( f \) is a real \( C^1 \) function defined on an open interval, and suppose that \( f(x_0) = y_0 \) and \( f'(x_0) \neq 0 \). Then there is an open interval about \( x_0 \) such that for every \( y \) sufficiently close to \( y_0 \) there is a unique solution \( x \) in this interval with \( f(x) = y \).

Proof: Let
\[ g_y(x) = x - f'(x_0)^{-1}(f(x) - y). \]  \hspace{1cm} (1.2)
This is set up so that \( g_{y_0}(x_0) = x_0 \). For each fixed \( y \) this has derivative \( g'_y(x) = 1 - f'(x_0)^{-1}f'(x) \) independent of \( y \). At \( x_0 \) this is \( g'_y(x_0) = 0 \). For \( |x - x_0| \leq r \) we have \( |g'_y(x)| \leq 1/2 \). Then by the mean value theorem
\[ |g_y(x') - g_y(x)| \leq (1/2)|x' - x|. \]  \hspace{1cm} (1.3)
Suppose \( |f'(x_0)^{-1}(y - y_0)| \leq r/2 \). Then \( g_y(x) - x_0 = g_{y_0}(x) - x_0 + f'(x_0)^{-1}(y - y_0) = g_{y_0}(x) - g_{y_0}(x_0) + f'(x_0)^{-1}(y - y_0) \) has magnitude at most \( (1/2)|x - x_0| + r/2 \leq r \). So \( |x - x_0| \leq r \) implies \( |g_y(x) - x_0| \leq r \). Thus we have a strict contraction on this closed interval to which the contraction mapping theorem applies.

This proof shows how one might compute an inverse function numerically. There is more to the inverse function theorem than this. It also says that the inverse function \( x = f^{-1}(y) \) is differentiable, and its derivative is \( dx/dy = 1/f'(f^{-1}(y)) \).

1.5 Compact metric spaces

A metric space \( M \) is compact if every sequence in \( M \) has a subsequence that converges to a point in \( M \).

If \( X \) is compact, and \( f : X \to Y \) is continuous, then \( f(X) \) is compact.

If \( X \) is compact, and \( f : X \to Y \) is continuous, then \( f : X \to Y \) is uniformly continuous.

Bolzano-Weierstrass. A metric space \( X \) is compact if and only if it is complete and totally bounded.

Proof in one direction: Suppose the space is totally bounded and complete. Consider a sequence \( n \mapsto s_n \) of points in \( X \).

Construct a sequence of balls \( B_k \) of radius \( 1/2^k \) and a decreasing sequence of infinite subsets \( N_k \) of \( N \) such that \( n \in N_k \) implies \( x_n \in B_k \).

This can be done as follows. Suppose that this has been done up to \( k - 1 \). Then the totally bounded space \( B_{k-1} \) can be covered by finitely many balls of radius \( 1/2^k \). There are infinitely many numbers in \( N_{k-1} \) and only finitely many
balls of radius $1/2^k$, so there must be one of the balls, call it $B_k$, and an infinite subset $N_k$ of $N_{k-1}$, such that $n \in N_k$ implies $x_n \in B_k$.

Pick $n_k$ strictly increasing with $n_k$ in $N_k$. This defines a subsequence that is Cauchy. By completeness it converges.

### 1.6 Hilbert space

Consider a vector space $H$ with an inner product $(u, v) \mapsto \langle u, v \rangle$. This has a norm $\|u\| = \sqrt{\langle u, u \rangle}$. If this normed space is a complete metric space, then $H$ is called a Hilbert space. A Hilbert space is a special kind of Banach space.

The main examples are $\mathbb{R}^n$ with the $\ell^2_n$ metric and $\ell^2$. The $\ell^2_n$ inner product is $\langle u, v \rangle = \sum_{i=1}^n u_i v_i$. The $\ell^2$ inner product is $\langle u, v \rangle = \sum_{i=1}^\infty u_i v_i$.

The main results are the theorem of Pythagoras, the Cauchy-Schwarz inequality, and the triangle inequality.

### 1.7 Appendix: matrix norms

Consider $m$ by $n$ real matrices $A$. There are various useful norms. Here are four of them.

1. $\|A\|_{1-1} = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|$. This has the property that $\|Ax\|_1 \leq \|A\|_{1-1} \|x\|_1$.

2. $\|A\|_{2-2} = \sigma_n$, where $\sigma_1^2 \leq \ldots \leq \sigma_n^2$ are the eigenvalues of $A^T A$. This has the property that $\|Ax\|_2 \leq \|A\|_{2-2} \|x\|_2$.

3. $\|A\|_{\infty-\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$. This has the property that $\|Ax\|_\infty \leq \|A\|_{\infty-\infty} \|x\|_\infty$.

4. $\|A\|_2 = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2}$.

The $\|A\|_{2-2}$ norm is difficult to compute. However there are two useful estimates.

1. $\|A\|_{2-2} \leq \|A\|_2$.

2. $\|A\|_{2-2} \leq \sqrt{\|A\|_{1-1} \|A\|_{\infty-\infty}}$. 
Chapter 2
Euclidean space and differentiable maps

2.1 Differentiable maps

Let $f$ be defined on an open subset of $\mathbb{R}^n$ with values in $\mathbb{R}^m$. We think of each of these as spaces of column vectors. Then $f$ may have a derivative $f'(x)$ at $x$ which is defined for $x$ in the open subset and has values that are $m$ by $n$ matrices. The defining property is the linear approximation

$$\frac{f(x+h)-[f(x)+f'(x)h]}{\|h\|} \to 0 \quad (2.1)$$

as $h \to 0$. We can also write this as

$$f(x+h) = f(x) + f'(x)h + \eta(x;h), \quad (2.2)$$

where $\eta(x;h)/\|h\| \to 0$ as $h \to 0$.

Notice that $f'(x)h$ is the product of the matrix $f'(x)$ with the column vector $h$ in $\mathbb{R}^n$. This product $f'(x)h$ is itself a column vector in $\mathbb{R}^m$.

The components of the matrix are written

$$f'(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}. \quad (2.3)$$

Thus

$$(f'(x)h)_i = \sum_j f'(x)_{ij}h_j. \quad (2.4)$$

If $y = f(x)$ we can also write

$$f'(x) = \frac{\partial y}{\partial x}. \quad (2.5)$$

This is a matrix equation.
The chain rule says that
\[ (g \circ f)'(x) = g'(f(x))f'(x). \] (2.6)

The product on the right is a matrix product.

The proof of the chain rule is from the definition of derivative. We have
\[ f(x + h) = f(x) + f'(x)h + \eta(x; h). \] Similarly, we have \( g(y + k) = g(y) + g'(y)k + \sigma(y; k). \) Take \( y = f(x) \) and \( k = f(x + h) - f(x). \) Then \( g(f(x + h)) = g(f(x)) + g'(f(x))(f(x + h) - f(x)) + \sigma(f(x), f(x + h) - f(x)). \) This may be written as
\[ g(f(x + h)) = g(f(x)) + g'(f(x))f'(x)h + g'(f(x))\eta(x, h) + \sigma(f(x), f'(x)h + \eta(x; h)). \] The problem is to show that the last two terms each go to zero even after being divided by the length \( \|h\|. \) This is more or less obvious for the first term. To deal with the second term we need to show that \( f'(x)h + \eta(x; h) \) not only goes to zero but in addition remains bounded when it is divided by \( \|h\|. \) But \( f'(x)h \) remains bounded when divided by \( \|h\|, \) since multiplication by a matrix is a Lipschitz function. The \( \eta(x; h) \) term divided by \( h \) even goes to zero, so in particular it remains bounded. The conclusion is that
\[ g(f(x + h)) = g(f(x)) + g'(f(x))f'(x)h + \tau(x; h), \] where \( \tau(x; h) \) divided by \( \|h\| \) goes to zero as \( h \to 0. \)

If \( z = g(y) \) and \( y = f(x) \) then this is the matrix equation
\[ \frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} \frac{\partial y}{\partial x}. \] (2.7)

The function \( f \) is said to be \( C^1 \) if \( f'(x) \) is continuous in \( f. \)

Mean value theorem. Say that \( f \) is a differentiable function. Suppose that \( a \) and \( b \) are points in \( \mathbb{R}^n \) and that for each \( x = (1 - t)a + tb \) on the segment between \( a \) and \( b \) we have
\[ \|f'(x)\| \leq M. \] (2.8)

Then
\[ \|f(b) - f(a)\| \leq M\|b - a\|. \] (2.9)

We can think of \( f'(x) \) as a function from \( x \) in the open subset of \( \mathbb{R}^n \) to the space of \( m \) by \( n \) matrices regarded as a Euclidean space of dimension \( mn. \) This has a derivative \( f''(x) \) with the property that
\[ \frac{f'(x + u) - [f'(x) + (f''(x))u]}{\|u\|} \to 0 \] (2.10)

The expression \( f''(x) \) has components \( f''(x)_{i,j,k}. \) The expression \( (f''(x))u \) has components \( (f''(x))_{i,j} = \sum_k f''(x)_{i,j,k}u_k. \) The expression \( (f''(x)h)u \) has components \( (f''(x)h)u_i = \sum_j \sum_k f''(x)_{i,j,k}h_ju_k. \) Sometimes the components are written
\[ f''(x)_{i,j,k} = \frac{\partial f_i(x)}{\partial x_k \partial x_j}. \] (2.11)

The function \( f \) is said to be \( C^1 \) if \( f'(x) \) is continuous in \( x. \) Similarly, it is \( C^2 \) if \( f''(x) \) is also continuous in \( x. \)
2.2. THE INVERSE FUNCTION THEOREM

Symmetry theorem. Suppose that \( f \) is a \( C^2 \) function. Then \( (f''(x)h)u = (f''(x)u)h \).

The symmetry theorem says that

\[
f''(x)_{ijk} = f''(x)_{ikj}
\]

or

\[
\frac{\partial f_i(x)}{\partial x_k \partial x_j} = \frac{\partial f_i(x)}{\partial x_j \partial x_k}.
\]

(2.12)

(2.13)

The idea of the proof of the symmetry theorem is to note that

\[
[f(x+h+u)−f(x+u)]−[f(x+h)−f(x)] = [f(x+u+h)−f(x+h)]−[f(x+u)−f(x)].
\]

(2.14)

The left hand side is an approximation to the derivative in the \( h \) direction followed by the derivative in the \( u \) direction, while the left hand side is an approximation to the derivative in the \( u \) direction followed by the derivative in the \( h \) direction.

2.2 The inverse function theorem

The context is a continuous function given by \( y = f(x) \) from a subset of \( \mathbb{R}^n \) to \( \mathbb{R}^n \).

Solving equations by iteration. Say that \( f \) is a continuous function defined on a ball \( B_r(x_0) \). Suppose that there is an invertible square matrix \( A \) such that \( g(x) = x - A^{-1}f(x) = A^{-1}(Ax - f(x)) \) is a contraction with Lipschitz constant \( 1/2 \). Suppose also that \( f(x_0) = y_0 \), and \( y \) satisfies \( |A^{-1}(y - y_0)| \leq r/2 \). Then \( g(x) = y \) has a unique solution in the interval.

Proof: Let

\[
g(x) = x - A^{-1}(f(x) - y).
\]

(2.15)

Then \( g \) is also Lipschitz with constant \( 1/2 \). Furthermore \( g(x) - x_0 = g(x) - g(x_0) + A^{-1}(y - y_0) \). This has magnitude at most \((1/2)|x - x_0| + r/2 \leq r \). So \(|x - x_0| \leq r \) implies \(|g(x) - x_0| \leq r \). Thus we have a strict contraction on this closed interval to which the contraction mapping theorem applies.

The context is a differentiable function given by \( y = f(x) \) from an open subset of \( \mathbb{R}^n \) to \( \mathbb{R}^n \). Then for each \( x \) the derivative \( f'(x) \) is a square \( n \) by \( n \) matrix. It has an inverse precisely when its determinant is zero.

Notice that if the function \( y = f(x) \) has an inverse function \( x = f^{-1}(y) \), and they are both differentiable, then from \( f(f^{-1}(y)) = y \) and the chain rule we get

\[
f'(f^{-1}(y))f^{-1'}(y) = I.
\]

(2.16)

Thus

\[
f^{-1'}(y) = f'(f^{-1}(y))^{-1}.
\]

(2.17)

Inverse function theorem. Suppose that \( y = f(x) \) defines a real \( C^1 \) function defined on an open subset, and suppose that \( f(a) = b \) and the matrix \( f'(a) \) is
invertible. Then there is an open subset \( U \) about \( a \) and an open subset \( V \) about \( b \) such that for every \( x \) in \( U \) the image \( f(x) \) is in \( V \), and such that for every \( y \) in \( V \) there is a unique solution \( x \) in \( U \) of \( f(x) = y \). Furthermore, the inverse function given by \( x = f^{-1}(y) \) is \( C^1 \) from \( V \) to \( U \).

Proof: Let
\[
g_y(x) = x - f'(a)^{-1}(f(x) - y). \tag{2.18}
\]
This is set up so that \( g_y(a) = a \). For each fixed \( y \) this has derivative \( g_y'(x) = I - f'(a)^{-1}f'(x) \) independent of \( y \). At \( a \) this is \( g_y'(a) = 0 \). For \( |x - a| \leq r \) we have \( |g_y'(x)| \leq 1/2 \). Then by the mean value theorem
\[
|g_y(x') - g_y(x)| \leq (1/2)|x' - x|. \tag{2.19}
\]
Suppose \( |f'(a)^{-1}(y - b)| \leq r/2 \). Then
\[
g_y(x) - a = g_y(x) - a + f'(a)^{-1}(y - b) = g_y(x) - g_y(a) + f'(a)^{-1}(y - b) \]
has magnitude at most \( (1/2)|x - a| + r/2 \leq r \). So \( |x - a| \leq r \) implies \( |g_y(x) - a| \leq r \). Thus we have a strict contraction on this closed ball to which the contraction mapping theorem applies. So for \( |f'(a)^{-1}(y - b)| \leq r/2 \) we have a solution of \( f(x) = y \) with \( |x - a| \leq r \).

Let \( V \) be the open subset of all \( y \) with \( |f'(a)^{-1}(y - b)| < r/2 \). Then in particular for each \( y \) in \( V \) we have a solution \( x \) of \( f(x) = y \). Let \( U \) be the subset consisting of all such solutions \( x \). Then \( f \) is a one-to-one function from the open subset \( U \) with \( x_0 \) in it to the open subset \( V \) with \( y_0 \) in it.

Now comes a headache. We need to show that the inverse function \( x = f^{-1}(y) \) is continuous. Consider a \( \delta > 0 \) so small that the closed ball of radius \( \delta \) about \( a \) is in \( U \), and also so that \( f'(x) \) is invertible for \( x \) in this ball. Then on this ball the function \( f \) has a continuous inverse \( f^{-1} \). This is because each closed subset \( S \) of the ball is compact, and so the inverse image of \( S \) under \( f^{-1} \), that is, the set \( f(S) \), is compact, hence closed.

Next consider the open ball of radius \( \delta \) about \( a \). Since \( f'(x) \) is invertible for each \( x \) in the ball, we can repeat the first argument at each point in the open ball to show that the image of this open ball is open. So we now have shown that \( f \) maps an open subset with \( a \) in it to an open subset with \( b \) in it in such a way that the inverse function is continuous.

With continuity of the inverse it is now possible to prove that the inverse is differentiable. Focus on a point \( a \) with \( f(a) = b \). Then \( f^{-1}(b) = a \). By continuity, if \( y \to b \), then \( f^{-1}(y) \to f^{-1}(b) \).

The trick is to write
\[
y = f(x) = f(a) + f'(a)(x - a) + \eta(a; x - a). \tag{2.20}
\]
This says that
\[
y - b = f'(a)(f^{-1}(y) - f^{-1}(b)) + \eta(a; f^{-1}(y) - f^{-1}(b)). \tag{2.21}
\]
In other words,
\[
f^{-1}(y) - f^{-1}(b) = f'(a)^{-1}(y - b) - f'(a)^{-1}\eta(a; f^{-1}(y) - f^{-1}(b)). \tag{2.22}
\]
Now use the continuity and this equation to prove that the inverse function is Lipschitz. As \( y \) approaches \( a \), the value \( f^{-1}(y) \) approaches \( f^{-1}(b) \). So \( \eta(a; f^{-1}(y) - f^{-1}(b)) \) divided by \( \|f^{-1}(y) - f^{-1}(b)\| \) approaches zero. In particular, if \( y \) is close enough to \( b \), then the last term is bounded by \( \frac{1}{2}\|f^{-1}(y) - f^{-1}(b)\| \). So we get

\[
\|f^{-1}(y) - f^{-1}(b)\| \leq \|f'(a)^{-1}(y - b)\| + \frac{1}{2}\|f^{-1}(y) - f^{-1}(b)\|. \tag{2.23}
\]

In other words,

\[
\|f^{-1}(y) - f^{-1}(b)\| \leq 2\|f'(a)^{-1}(y - b)\| \leq 2\|f'(a)^{-1}\||y - b|. \tag{2.24}
\]

With this out of the way, we can get differentiability. Let \( y \to b \). Then by continuity \( f^{-1}(y) - f^{-1}(b) \to 0 \). Furthermore the \( \eta \) vector divided by the norm of this vector approaches zero. However we want instead to divide by the norm of \( y - b \). To compensate, we need a bound on the ration of the two norms. From the Lipschitz argument we see that this ratio is indeed bounded by \( 2\|f'(a)^{-1}\| \).

We have shown that \( x = f^{-1}(y) \) from \( V \) to \( U \) is differentiable. The formula for the derivative is that given before:

\[
f^{-1}'(y) = f'(x)^{-1} = f'(f^{-1}(y))^{-1}. \tag{2.25}
\]

This in particular shows that the inverse function is \( C^1 \).

### 2.3 The implicit function theorem

**Theorem.** Say that \( f(x, u) \) is a \( C^1 \) function and \( f(x_0, u_0) = c \) and \( \partial f(x_0, u_0)/\partial u \) is invertible. Then for \( x \) near to \( x_0 \) there is a unique \( C^1 \) function \( g(x) \) with \( g(x_0) = u_0 \) and \( f(x, g(x)) = c \).

**Proof:** Define \( h(x, u) = (x, f(x, u)) \). We have \( h(x_0, u_0) = (x_0, c) \). Apply the inverse function theorem to \( h \). The components of the inverse function function are \( m(z, w) \) and \( n(z, w) \), and they satisfy \( m(z, w) = z \) and \( f(m(z, w), n(z, w)) = w \). Furthermore, they are defined uniquely near \( z = x_0 \) and \( w = c \) and satisfy \( m(x_0, c) = x_0 \) and \( n(x_0, c) = u_0 \). This implies in particular that \( f(x, n(x, c)) = c \). Take \( g(x) = n(x, c) \). Then \( g(x_0) = u_0 \), and \( f(x, g(x)) = c \), as desired.