Analysis 2 - Thomas Mountford

Alp Ozen

Spring 2020

Contents

1   Notions on $\mathbf{R}^n$

1.1   Introducing topological properties on $\mathbf{R}^n$

Let's recall that $\mathbb{R}^n$ is a Euclidean vector space. We define a scalar product on $\mathbb{R}^n$ as follows:

Definition 1.

$$< \cdot, \cdot >: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

1. $< x, x > \geq 0$

2. $< x, y > = < y, x >$

3. $< ax + by, z > = a < x, z > + b < y, z >$

A *norm* is defined as function that maps some real vector space E to $\mathbb{R}$ and satisfies:

$$1)\ |x| \geq 0\ \forall x \in E,\ |x| = 0 \iff x = 0$$
$$2)\ |\lambda \cdot x| = |\lambda| \cdot |x|$$
$$3)\ |x + y| \leq |x| + |y|$$

In our intuitive understanding of $\mathbb{R}^n$ we are actually thinking about the Euclidian space $\mathbb{R}^n$ equipped with the Euclidian norm.

Definition 2. *Euclidian norm*

$$|x|_2 = \sqrt{< x, x >} = \left(\sum_i^n x_k^2\right)^{\frac{1}{2}}$$

And from this naturally follows the definition of Euclidian distance:

Definition 3.

$$d(x, y) = |x - y|$$

We note that d satisfies the same 3 properties as the norm. Thus, the couple (E, d) is called a metric space.

And now more definitions:

Definition 4. *Open sets*

1. **Open ball** $B(a, r) := \{x \in \mathbb{R}^n : d(x, a) < r\}$

2. **Open subset** *Some subset $S \subset \mathbb{R}^n$ is open if $\forall x \in \mathbb{R}^n$, $\exists \epsilon > 0\ B(x, \epsilon) \subset S$*

3. **Closed subset** *Some S is closed if $\mathbb{R}^n - S$ is open, note that the empty set and $\mathbb{R}^n$ are both open and closed.*

4. **The interior and boundary of a set** *a is in the interior of S if $\exists \epsilon > 0\ B(a, \epsilon) \subset S$ and b is in the boundary of a set S if any $B(a, \epsilon)$ contains points from both S and $\mathbb{R}^n - S$. The set of all interior points is denoted $(^\circ S)$ and set of all boundary points is denoted ∂S*

5. **Closure of a set** *a is a closure of S if for any $B(a, \epsilon)$ we have $B(a, \epsilon) \cap S \neq \emptyset$*

Definition 5. *Topology*
  A Topology exists whenever the following are satisfied:

$$\text{For a given } M \subset \mathbb{R}^n \text{we define } O \subset P(M)$$

1. $\emptyset \in O$, $M \in O$

2. $U \in O, V \in O \to U \cap V \in O$

3. $U_\alpha \in O \to \bigcup_\alpha U_\alpha \in O$

Definition 6. ***Closure of a set*** *A point $a \in \mathbb{R}^n$ is a closure point of S if for any $B(a, \epsilon)$ we have:*

$$B(a, \epsilon) \cap S \neq \emptyset$$

The set of all closure points called the closure of S is denoted $\bar{S}$ noting that this only holds in $\mathbb{R}^n$

$$\bar{S} = S \cup \partial S$$

Theorem 1.1. *Important results on closures and boundaries*

$$S^\text{o} \subset S \subset \bar{S}$$
$$\bar{S} = S^\text{o} \cup \partial S$$
$$S \text{ is open iff } S = S^\text{o}$$
$$S \text{ is closed iff } S = \bar{S}$$

Definition 7. ***Sequence in $\mathbf{R}^n$***

$f : \mathbb{N} \to \mathbb{R}^n$ such that:
$$x_k = f(k) \in \mathbb{R}^n \forall k \in \mathbb{N}$$

Definition 8. ***Convergence***

The definition of convergence in $\mathbb{R}^n$ is very similar to its counterpart in $\mathbb{R}$. We say that x_k converges to x iff:

$$\forall \epsilon > 0 \ \exists N_\epsilon \ \forall k \geq N_\epsilon \ d_2(x_k, x) < \epsilon$$

Definition 9. ***Complete spaces aka. Banach spaces***
A space is called complete if every Cauchy sequence in this space converges to a limit. Some example of complete spaces are $\mathbb{R}, \mathbb{R}^n$

We restate the axioms of a **Norm** and of a **Metric**. The subtle difference between the two is that a norm can only be applied to some vector space whereas a metric is applicable to other spaces.

Remark 1.1. ***Axioms of norm and metric***
Norm:

1. $||x|| \geq 0$

2. $||\lambda x|| = |\lambda| ||x||$

3. $||x + y|| \leq ||x|| + ||y||$

 Metric:

1. $d(x, y) \geq 0$

2. $d(x, y) = d(y, x)$

3. $d(x, y) \leq d(x, z) + d(z, y)$

1.2   Difficult problems and notes from week 2

We note that any open set is equal to its interior.

Example 1.1. *Consider the set defined as $T = \{(x,y) \in \mathbb{R}^2 | 1 < x^2 + 4y^2 < 4\}$ We observe the following about this set:*

$$T^{\circ} = T$$
$$\partial T = \{(x,y) \in \mathbb{R}^2 | 1 = x^2 + 4y^2 \, or \, x^2 + 4y^2 = 4\}$$

Remark 1.2. *Rather interestingly, the rational numbers are neither closed nor open because both the irrationals are dense as well as the rationals being dense which was proved in analysis 1. We have the following properties for $\mathbb{Q}$*

$$\partial \mathbb{Q} = \bar{\mathbb{Q}} = \mathbb{R}$$

Definition 10. *If x_k is in a bounded closed set D then $\exists$ a convergent subsequence with limit in D.*

Definition 11. *Continuity open ball definition*
$\forall a \in \mathbb{R}^n$ we have $\forall \epsilon > 0 \, \exists \delta > 0$ such that $f(B(a,\delta)) \subset B(f(a),\epsilon)$

Definition 12. *Continuity open set definition*
Let $f : \mathbb{R}^n \to \mathbb{R}^m$ Then f is continuous on $\mathbb{R}^n$ iff. for all $O \subset \mathbb{R}^m$ where O is open, we have $f^{-1}(O)$ open in $\mathbb{R}^n$

Proof. **(only if part)** Let f be continuous, we want to show that whenever O is open, $f^{-1}(O)$ is also open. Now take $f^{-1}(O)$. Let $p \in f^{-1}(O)$. Then $f(p) \in O$ Now we have that O is open hence $\exists \epsilon > 0$ s.t. $B(f(p),\epsilon) \subset O$ Now because f is continuous at p we have that $\exists \delta > 0$ s.t. $\underbrace{f(B(p,\delta)) \subseteq B(f(p),\epsilon)}_{\text{applying open ball cont.}} \subseteq O$

This further shows that $B(p,\delta) \subseteq f^{-1}(O)$ which shows that f is open in S.

(if part) Now suppose that for all open sets $O \in \mathbb{R}^m$ we have $f^{-1}(V)$ are open in S. To show that f is continuous, we must show f is continuous on all $p \in S$. Let $p \in S$ s.t. $y = f(p)$. Then $\forall \epsilon > 0$ $B(y,\epsilon)$ is open. By assumption then, $f^{-1}(B(f(p),\epsilon))$ is open in S. Now, $f(p) \in B(f(p),\epsilon)$ we have $p \in f^{-1}(B(f(p),\epsilon))$ is open in S, $\exists \delta > 0$ s.t. $B(p,\delta) \subseteq f^{-1}(B(f(p),\epsilon))$ and applying f to both sides we get our result. $\square$

1.3   More norms on $\mathbf{R}^n$

Consider the non-usual norms $|x|_1 := \sum |x|$ and $|x|_\infty = max|x_k|$. A super interesting result is that all norms on $\mathbb{R}^n$ are equivalent.

Definition 13. *Two norms $|\cdot|$ and $||\cdot||$ are **equivalent** if $\exists C_1, C_2 > 0$ such that $\forall x \in E$:*

$$C_1|x| \leq ||x|| \leq C_2|\cdot|$$

Theorem 1.2. *Any two norms on $\mathbb{R}^n$ are equivalent.*

We do not prove this theorem but consider the visual below for why it is true.
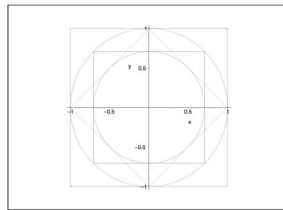

Figure 1: Norm equivalence

1.4   Curves in $\mathbf{R}^n$

Definition 14. *Curve* *A curve is a function $f : I \to \mathbb{R}^n$ denoted as:*

$$f = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$$

A good example of a curve is the speed of a particle in mechanics. It takes only one input t and has a different function defined for each coordinate. Another example is the graph $G_f = \begin{bmatrix} x \\ f(X) \end{bmatrix}$ of any function from $I \to \mathbb{R}^k$ Another curve is a helix defined as:

Definition 15. *Helix*

$$f(t) = \begin{bmatrix} r \cos t \\ r \sin t \\ ct \end{bmatrix}$$
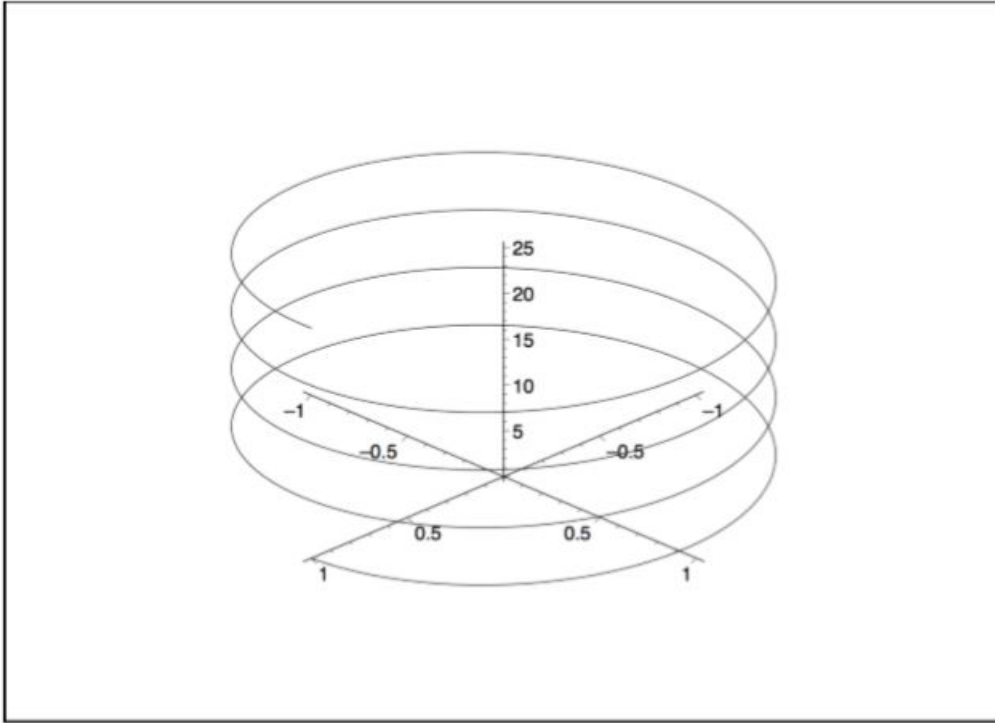
Figure 2: Helix

Definition 16. *Arc-length of a curve* *Given a curve is rectifiable(each segment tends to 0 and sum of segments tend to $L > 0$ we define*

$$L = \int_b^a |f'(t)|_2 dt$$

Now the above formula is taught at high school level as:

$$L = \int_b^a f' \sqrt{1 + \frac{dx}{dy}^2}\, dx$$

which is simply the application of our definition to $\mathbb{R}^2$

1.5   Real-valued functions in $\mathbf{R}^n$

Definition 17. _Real valued function_ $f : D \to \mathbb{R}$ _such that_ $D \subset \mathbb{R}^n$

The graph G_f of a real valued function is the set $G_f = \{\begin{bmatrix} x \\ f(x) \end{bmatrix} || x \in D\} \subset \mathbb{R}^{n+1}$

Hence where a curve is map from one dimension to n dimensions, a real-valued function is a map from n dimensions to 1 dimension.

1.6   Partial derivatives

Definition 18. $f : \mathbb{R}^n \to \mathbb{R}$ _is partially differentiable with respect to variable_ x_k _if:_

$$\frac{\partial f}{\partial x_k}(a) = \lim \frac{f(a + he_k) - f(a)}{h}$$

exists. To calculate the partial derivative at x_k _we simply let all_ x_i _be constants._

Definition 19. _Gradient_ _The gradient of_ f _is defined as:_

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{bmatrix} = \sum_{k=1}^{n} D_k f(x) e_k$$

Example 1.2. _Let_ $f(x) = |x|_2$ _Then we have that_ $\frac{\partial f}{\partial x_k}(a) = \frac{x_k}{r}$

Now we come to a more complicated example:

Example 1.3.

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2}(x,y) \neq 0 \\ 0(x,y) = 0 \end{cases}$$

In this case, we have that although the partial derivative exists its existence does not imply continuity. The partial derivative itself must be continuous to imply continuity.

Definition 20. _Continuity in_ $\mathbf{R}^n$

We say that $f : \mathbb{R}^n \to \mathbb{R}$ _is continuous at_ $a \in \mathbb{R}^n$ _if_ $\forall \epsilon > 0 \ \exists \delta > 0$ _such that_ $|x - a|_2 < \delta \to |f(x) - f(a)|_2 < \epsilon$

The problem with the above definition is that we may approach say $(0, 0)$ from infinite number of ways(easily seen considering a cartesian plane). This was not the case in $\mathbb{R}^{\Vdash}$ because we could approach 0 from the left or the right. Hence to prove that a limit exists by using cases is impossible, we need to instead use the epsilon-delta definition.

Definition 21. _We define_ $C^1(U)$ _as the set of functions_ $f : U \subset \mathbb{R}^n \to R$ _such that_ f _has_ n _partial derivatives each continuous._

Definition 22. _Contour graphs_ _When we are presented with a real-valued function of multiple variables, we may either represent it via its graph, that is map it to_ $\mathbb{R}^{m+1}$ _or instead draw its contour lines that is we let the function equal some value_ $c \in \mathbb{R}$ _and plot all the points in_ $\mathbb{R}^{\triangleright}$ _that take this value._

And now we define the total differential.

Definition 23. _Total differential_

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy$$

Well when might the total differential ever be useful? Here's a simple example.

Example 1.4. *Let's define $z = \sqrt{x^2 + y^2}$ Now suppose we want to evaluate the function at $(x = 3.92, y = 4.01)$ Then we may simply evaluate the function at $(x = 4, y = 4)$ and using the total differential find the change in z and subtract it from our initial result.*

Definition 24. *__Bilinear form__ We may think of a bilinear form as a linear map that accepts 2 arguments instead of one and linearity holds for both arguments.*

$$B : V \times V \to F$$

$$B(v_1 + v_2, w) = B(v_1, w) + B(v_2, w)$$
$$B(fv, w) = fB(v, w)$$
$$B(v, w_1 + w_2) = B(v, w_1) + B(v, w_2)$$
$$B(v, fw) = fB(v, w)$$

We now define differentiability using the total differential.

Definition 25. *A function f is said to be differentiable in $a \in U \subset \mathbb{R}^n$ if $\boldsymbol{\nabla} f(a)$ exists and:*

$$\lim_{h \to 0} \frac{f(a + h) - f(a) - \underbrace{< \boldsymbol{\nabla} f(a), h >}_{total\ differential}}{h} = 0$$

The above definition assures that the total change in f over some variable h occurs quicker than the change in h.

Similarly, we define directional derivatives.

Definition 26. *For some $f : U \subset \mathbb{R}^n \to \mathbb{R}$ the directional derivative at a with respect to $v \in U$ is:*

$$D_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t}$$

Now some non-implications to pay attention to.

Remark 1.3. *Non-implications*

- *If all directional derivatives at a exist, it does not mean that f is differentiable at a.*

- *Directional derivatives with respect to v are not neccesarily equal to $< \boldsymbol{\nabla} f, v >$ this is only true when f is differentiable at a.*

Definition 27. *__Notation for higher order derivatives__ We notate a derivative of order 3 with respect to x,y,x in respective order as:*

$$D_{x,y} f(x)$$

Definition 28. *__Tangent hyperplane equation__ For some $f : U \subset \mathbb{R}^n \to \mathbb{R}$ we define the tangent hyperplane at some point $a \in U$ as:*

$$f(\bar{x}) = f(a) + < \boldsymbol{\nabla} f(a), h >$$

Definition 29. *__Generalized chain rule__ Let k be a curve from $\mathbb{R} \to \mathbb{R}^n$ and f a function from $\mathbb{R}^n \to \mathbb{R}$ such that:*
1. k is differentiable at $t_0 \in \mathbb{R}$
2. f is differentiable at $k(t_0)$
    Then we have that:

$$(fok)(t_0)' = < \boldsymbol{\nabla} f(k(t_0)), k'(t_0 >$$

Proof. We briefly prove the above theorem here.

First let's note the following: The fact that k is differentiable at t_0 means that

$$k(t_0 + h) - k(t_0) = k'(t_0)h + o(|h|)$$

From this it follows that the LHS is also $O(|h|)$ hence if some $s(h)$ is little o of the LHS, it is definitely little of the RHS. Now we have:

$$f(k(t_0 + h) - k(t_0)) - f(k_0) - < \boldsymbol{\nabla} f(k(t_0)), k(t_0 + h) - k(t_0) > +o(|k(t_0 + h) - k(t_0)|) = 0$$
$$f(k(t_0 + h) - k(t_0)) - f(k_0) = h < \boldsymbol{\nabla} f(k(t_0)), k'(t_0) > +o(|h|)$$

$\square$

Now we define higher order derivatives.

Definition 30. *A partial derivative of second order is written as:*

$$\frac{\partial^2 f}{\partial x_j \partial x_i} \equiv D_{ji}f = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_i}(\bar{a} + he_j) - \frac{\partial f}{\partial x_i}(\bar{a})}{h}$$

An interesting result is that the order of partial differentiation does not matter.

Theorem 1.3. *If $f \in C^m(U)$ then $\forall r \leq m$ we have that:*

$$D_{j_1,\ldots,j_r} = D_{i_1,\ldots,i_r}$$

where the i are a permutation of the j

Now we introduce the hessian.

Definition 31. ***Hessian matrix***

$$Hess_f(x) = \begin{bmatrix} D_{11} & \ldots & D_{1n} \\ \vdots & \ddots & \vdots \\ D_{n1} & \ldots & D_{nn} \end{bmatrix}$$

Using the hessian, we define the second order taylor approximaion of a multivariable functions as below.

Theorem 1.4.
$$f(\bar{x} + \bar{h}) = f(\bar{x}) + < \boldsymbol{\nabla} f(\bar{x}), \bar{h} > +\frac{1}{2}\bar{h}^T Hess(\bar{x})h$$

Intuition for the above theorem is easy. Just remember that the $\frac{1}{2}$ comes from the fact that it is a taylor expansion (analogue in 1-d is $f(x) = f(a) + f'(a)(x - a) + \frac{1}{2}f''(x - a)^2$)

Theorem 1.5. *Whenever we have a product, the following holds:*

$$\boldsymbol{\nabla}(fg) = \boldsymbol{\nabla} fg + \boldsymbol{\nabla} gf$$

1.7   Vector Fields

Definition 32. ***Vector field*** *A vector field is simply a function from $\mathbb{R}^n \to \mathbb{R}^m$*

How might we imagine the partial derivative of a vector field?

Definition 33. *Some $v : U \to \mathbb{R}^m$ is partially differentiable at $\bar{a} \in U$ if:*

$$D_j v_i(a) = \frac{\partial v_i}{\partial x_j}(a) \equiv \lim_{h \to 0} \frac{v_i(a + he_j) - v_i(a)}{h} \ exists$$

We introduce the Jacobian matrix as follows:

Definition 34. *Jacobian*

$$J_f(x) = \begin{bmatrix} D_1 v_1 & \dots & D_n v_1 \\ \vdots & \ddots & \vdots \\ D_1 v_n & \dots & D_n v_n \end{bmatrix}$$

Given this definition of the Jacobian, we define differentiability at some point $a \in U \subset \mathbb{R}^n$ as:

Definition 35. *Differentiability Some $v : U \to \mathbb{R}^m$ is differentiable at $a \in U$ if there exists a linear map J_f such that*

$$\lim_{h \to 0} \frac{v(a+h) - v(a) - J_f(h)}{||h||} = 0$$

Now the same concept of $C^1(\mathbb{R})$ is applicable to vector fields. Some vector field v is of class $C^1(U$ if all partial derivatives $D_j v_i : U \to \mathbb{R}$ exist and are continuous in U.

And now we ask what the Jacobian might be for the composition of two vector spaces?

Theorem 1.6. *The composition rule is exactly like the chain rule in single variable calculus. For some $v : \mathbb{R}^n \to \mathbb{R}^m$ and some $w : \mathbb{R}^m \to \mathbb{R}^n$ we have*

$$J_{w \ o \ v}(x) = J_w(v(x)) J_v(x)$$

Hence an intuitive way to calculate the i, j position element of $J_{w \ o \ v}(x)$ is to calculate:

$$\frac{\partial w_i(v(x))}{\partial x_j} = \sum_{k=1}^{n} \frac{\partial v_k}{\partial x_j}(x) \frac{\partial w_i}{\partial v_k} v(x)$$

Let's build a better intuition for the Jacobian matrix. It's columns are simply partial derivatives of $v(x)$ with respect to x_i. Or conversely, the rows of the Jacobian are the gradient of *ith* row of the function. Hence we have with i fixed a constant:

$$J_W(x)_{i,j} = \boldsymbol{\nabla} f(x)_i^T$$

1.8 Extrema

For some critical point a, we have that $\boldsymbol{\nabla} f(a) = 0$. This means that locally our function when approximated at second order becomes:

$$f(x) = f(a) + (x-a)^T Hessf(a)(x-a)$$

. Now if all eigenvalues of $Hessf(a)$ are positive then no matter which direction we go, well have that a is a minimum. Similarly, if all eigenvalues of $Hessf(a)$ are negative, then a is a minimum. Finally, if some are negative and some positive, we have a saddle point. To stress why this is true, just realize that an eigenvalue is defined as $Hessf(a) = \lambda a$

1.9 Implicit function theorem

Consider the real-valued function $f(x,y) = x^2 + e^y - 1$ Now our goal is to determine the level set $N_f(0)$. Hence we write $x^2 + e^y - 1 = 0$ and now our goal is to find y as a function of x. Well as it turns out $y = g(x) = \ln 1 - x^2$. Now $g'(x) = \frac{-2x}{1-x^2}$ which also happens to be $g'(x) = \frac{-D_1 f(x,g(x))}{D_2 f(x,g(x))}$ Hence for some $U \subset \mathbb{R}^{\not{k}}$ we state the following:

Theorem 1.7. *Let $f : U \to \mathbb{R}$ be of class C^1 (ie. continuous partial derivatives) such that $f(a,b) = 0$ and also $D_2 f(a,b) \neq 0$ Then $\exists B(a,\epsilon)$ and a unique function $g : B(a,\epsilon) \to \mathbb{R}$ such that:*
1. graph of g is in U
2. $g(a) = b$
3. $f(x,g(x)) = 0$
4. $g'(x) = \frac{-D_1 f(x,g(x))}{D_2 f(x,g(x))}$

We may now also generalize to higher dimension:

Theorem 1.8. *Implicit function theorem* *Let $F \in C^1$ such that $F(\bar{x}, y) = 0$ and $\frac{\partial F}{\partial y} \neq 0$ Then there exists a unique $y = f(\bar{x})$ such that*
1. graph of f is in U
2. $f(\bar{x}) = y$
3. $F(\bar{x}, y) = 0$
4. $\frac{\partial f}{\partial x_i} = \frac{-\frac{\partial F}{\partial x_i}}{\frac{\partial F}{\partial y}}$

In simpler words, the partial of $g(x)$ relative to the *jth* variable is the partial of F relative to the *jth* variable divided by the partial of F relative to y.

Now let's put our above theorem to use. Let the level surface $F(x_1, \ldots, x_n) = 0$ contain $\bar{a}$ and let $\frac{\partial F}{\partial x_n}(\bar{a}) \neq 0$ Hence we may write:

$$x_n = g(x_1, \ldots, x_{n-1})$$

Hence locally, the surface F is like the $n-1$ dimensional function g. Thus we may write:

$$x_n - a_n = \sum_{j=1}^{n-1} \frac{\partial g}{\partial x_j}(a_1, \ldots, a_{n-1})(x_j - a_j)$$

But given our earlier theorem this simplifies to:

$$\sum_{j=1}^{n} D_j F(\bar{a})(x_j - a_j) = 0$$

Before ending this chapter, I'd like to make a remark about gradients. For a multivariable function, our intuition of a gradient from 1-d no longer holds. Now we define a gradient as a linear map from $\mathbb{R}^n \to \mathbb{R}$ as $D_{F,p}(\langle a, b \rangle) = [\frac{\partial F}{\partial x} \frac{\partial F}{\partial y}] \begin{bmatrix} a \\ b \end{bmatrix}$. Now suppose we are at some point a on the level surface. Then by definition the tanget direction vector h will be the vector such that $D_{F,p}(h) = 0$. Using the linear map form this is equivalent to saying $\langle \frac{\partial F}{\partial x} \frac{\partial F}{\partial y} \rangle \cdot \langle h_1, h_2 \rangle = 0$ which is equivalent to saying that the gradient at some point is perpendicular to that level surface.

Finally we also note that whenever some vector field is invertible, the derivative of the inverse field is equal to the inverse of the Jaacobian. Finally we also make an important distinction below:

Definition 36. *Local vs global invertibility* *Local invertible means invertible on the linear approximation which is the Jacobian, globally invertible means the inverse in the traditional sense.*

1.10   Optimization

We now define critical points and extrema for multivariable functions.

Definition 37. *Extrema*

- *$a \in U$ is a local extremum of f if $\exists \delta$ such that $f(a) \geq$ or $\leq f(x) \forall x \in B(a, \delta)$*

- *$a \in U$ is a critical/stationary point of f if $\boldsymbol{\nabla} f(a) = 0$*

- $a \in U$ *is a saddle point if* $\boldsymbol{\nabla} f(a) = 0$ *but a is not a local extremum*

Some practical points an engineer must remember about extrema are that the Hessian is extremely useful in analyzing extrema, that is:

- If we have a critical point and also the eigenvalues of the Hessian are positive-definite, than f admits a local minimum(resp. local max when negative-definite)

- If we have a critical point and the eigenvalues of the Hessian are mixed, than we have a saddle point.

Yet another test for local max./min. and saddle point is the following:taken from stackOverflow answer of user/169780

$det(Hf(a)) = 0$ means that there is a zero eigenvalue and so the test fails.
$det(Hf(a)) < 0$ means that both eigenvalues have different sign, so we have a saddle point at a.
$det(Hf(a)) > 0$ means that both eigenvalues have the same sign: either both positive or both negative, and we must use the trace to decide which it is. In fact, rather than use the trace, it actually suffices to just use the top left entry $\frac{\partial^2 f}{\partial x^2}(a)$ of $Hf(a)$ to decide, by Sylvester's criterion. In other words, $\frac{\partial^2 f}{\partial x^2}(a) > 0$ means both eigenvalues are positive (local min at a a), whereas $\frac{\partial^2 f}{\partial x^2}(a) < 0$ means both eigenvalues are negative (local max at a).

We now naturally ask how we go about finding a min/max given a constraint on our variables or the domain we can consider. For instance consider the graphic below:
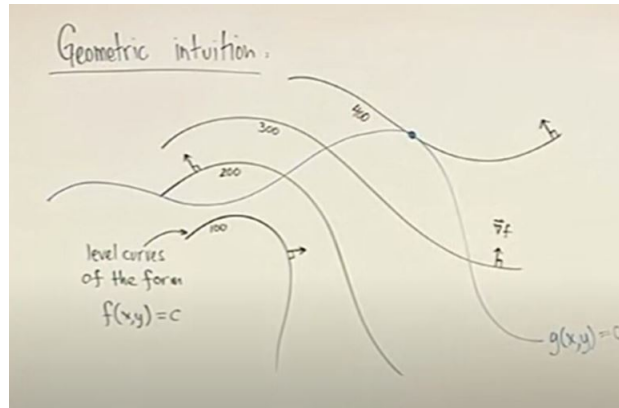
We have limited our x, y for some two variable f to equal 0. We then ask, with this constraint, at which x, y does f attain a maximum or a minimum. This is solved by lagrange multipliers

Theorem 1.9. *Let* $h, f : U \to \mathbb{R}$ *be of class* C^1 *and let* $M = \{x \in U | f(x) = 0\}$ *Now let* $a \in M$ *and let a be such that* $h(a)$ *is a local extremum. Then* $\exists \lambda :$

$$\boldsymbol{\nabla} h(a) + \lambda \boldsymbol{\nabla} f(a) = 0$$

1.11   Multiple integrals

Define a function

$$g(y) = \int_a^b f(x, y) dx$$

Now if f is continuous and its partial derivatives are continuous, f is $C^1(I)$ and we have:

$$g'(y) = \int_a^b \frac{\partial f(x, y)}{\partial y} dx$$

Using this property, our goal is to perform integrating through differentiation. Thus we have that over a fixed finite interval, if g is continuous, the integral is continuous. Similarly if g is continuously differentiable, then the integral is continuously differentiable.

Example 1.5. *Suppose we want to evaluate $\int_0^b x^2 \cos x dx$. We define $g(y) = \int_0^b \cos xy dx$ Now by our above theorem we have $g' = -\int_0^b x \sin xy dx$ and $g'' = -\int_0^b x^2 \cos xy dx$. And now $-g''(1) = \int_0^b x^2 \cos x dx$ Now we may simply evaluate the integral by evaluating g and differentiating twice.*

A critical result is **Fubini's theorem**:

Theorem 1.10. *__Fubini's theorem__ Let $f : D \to \mathbb{R}$ How may we then define the following double integral*

$$\int \int_D f(x,y) dx dy$$

As it turns out, if we have some:

$$g(x) = \int_a^b f(x,y) dx \quad h(x) = \int_d^c f(x,y) dy$$

with $D = [a,b] \times [c,d]$ then it holds that:

$$\int \int_D f(x,y) dx dy = \int_c^d g(y) dy = \int_a^b h(y) dx$$

1.12 Change of variables

Let's first of all present the change of variables formula for 2-d:

Theorem 1.11. *__Change of variables in 2-d__*

$$\int \int_R f(x,y) dy dx = \int \int_G f(g(u,v), h(u,v)) J(u,v) du dv$$

where $J(u,v)$ known as the Jacobian is:

$$J(u,v) = \det \left\{ \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial x}{\partial v} \end{bmatrix} \right\}$$