Variational Calculus - Week 0 - Prequel Lectures: Several Variable Calculus and Topology

Antonio León Villares

September 2022

Contents

Based on the notes by Jelle Hartong, Appendix A

1 Analysis & Algebra Recap

1.1 The Dot Product

- **What is the dot product?**
  - a **positive-definite** inner product in Euclidean space, $\mathbb{R}^n$:
  $$\langle -, - \rangle : \mathbb{R}^n \to \mathbb{R}$$
  - defined by:
  $$\langle \underline{x}, \underline{y} \rangle = \sum_{i=1}^{n} x^i y^i$$
  - here I am using the convention of the notes, whereby x^i denotes the ith component of the vector $\underline{x} \in \mathbb{R}^n$
- **What are the properties of the dot product?**
  - **Symmetry**:
  $$\langle \underline{x}, \underline{y} \rangle = \langle \underline{y}, \underline{x} \rangle$$
  - **(Sesqui) Linearity**:
  $$\langle \lambda \underline{x}, \underline{y} \rangle = \lambda \langle \underline{x}, \underline{y} \rangle, \qquad \lambda \in \mathbb{R}$$
  - **Positive Definite**:
  $$\langle \underline{x}, \underline{x} \rangle \geq 0$$
  with equality **if and only if**:
  $$\underline{x} = \underline{0}$$
- **What is the norm in Euclidean space?**
  - **geometrically**, the **distance** between a vector and the origin
  - defined via the dot product:
  $$\|\underline{x}\| = \sqrt{\langle \underline{x}, \underline{x} \rangle} = \sqrt{\sum_{i=1}^{n} (x^i)^2}$$
- **What 2 inequalities are satisfied by the dot product/norm?**
  - **Triangle Inequality**:
  $$\|\underline{x} + \underline{y}\| \leq \|\underline{x}\| + \|\underline{y}\|$$
  - **Cauchy-Schwarz Inequality**:
  $$|\langle \underline{x}, \underline{y} \rangle| = \|\underline{x}\| \|\underline{y}\|$$
- **What is an open ball?**

1.2 Topological Balls

- **What is an open ball?**
  - an **open ball of radius** r **centered at** $\underline{x}$ is the set:

$$B_r(\underline{x}) = \{\underline{y} \mid \underline{y} \in \mathbb{R}^n,\ \|\underline{x} - \underline{y}\| < r\}$$

- **What is a closed ball?**
  - a **closed ball of radius** r **centered at** $\underline{x}$ is the set:

$$B_r(\underline{x}) = \{\underline{y} \mid \underline{y} \in \mathbb{R}^n,\ \|\underline{x} - \underline{y}\| \leq r\}$$

- **What is an interior point?**
  - let $U \subset \mathbb{R}^n$, and consider a point $\underline{x} \in U$
  - $\underline{\text{x}}$ is an **interior point** of U if there exists an open ball, centered at $\underline{x}$, completely encompassed within U:

$$\exists \varepsilon > 0 \ : \ B_\varepsilon(\underline{x}) \subset U$$

- **What is an open subset?**
  - $U \subset \mathbb{R}^n$ is **open** if **every** $\underline{x} \in U$ is an **interior point of** U

1.3 Continuity

1.3.1 Definition: Continuous Function

*Let U be an **open subset** of $\mathbb{R}^n$ and define:*

$$f : U \to \mathbb{R}^m$$

*f is **continuous** at $\underline{a} \in U$ if:*

$$\forall \varepsilon > 0, \exists \delta > 0 \ : \ \|\underline{x} - \underline{a}\| < \delta \implies \|f(\underline{x}) - f(\underline{a})\| < \varepsilon$$

*f is **continuous** if it is continuous $\forall a \in U$.*

1.3.2 Definition: Continuous Function (via Topological Balls)

*Let U be an **open subset** of $\mathbb{R}^n$ and define:*

$$f : U \to \mathbb{R}^m$$

*f is **continuous** at $\underline{a} \in U$ if:*

$$\forall \varepsilon > 0, \exists \delta > 0 \ : \ \underline{x} \in B_\delta(\underline{a}) \implies f(\underline{x}) \in B_\varepsilon(f(\underline{a}))$$

[Equation A.5]

1.3.3 Theorem: Topological Characterisation of Continuity

The following gives an ***equivalent*** *definition of continuity to the* $\varepsilon - \delta$ *definition.*

> *Let* U *be an* ***open subset*** *of* $\mathbb{R}^n$ *and define:*
>
> $$f : U \to \mathbb{R}^m$$
>
> f *is* ***continuous*** *if and only if for any* ***open subset*** $V \subset \mathbb{R}^m$*, there exists an* ***open subset*** $W \subset \mathbb{R}^n$ *with:*
>
> $$f^{-1}(V) = W \cap U$$
>
> *where:*
>
> $$f^{-1}(V) = \{\underline{x} \mid \underline{x} \in U,\ f(\underline{x}) \in V\}$$
>
> *[Equation A.7]*

2 Derivatives and Change of Coordinates

[For this, really recommend Stewart's Calculus: Early Transcendentals.]

2.1 The Directional Derivative

- **What is a scalar field?**
  - a mapping from a vector to a scalar:
    $$f : \mathbb{R}^n \to \mathbb{R}$$

2.1.1 Definition: The Directional Derivative of a Scalar Field

> *Let* $U \subset \mathbb{R}^n$ *and define the* ***scalar field****:*
>
> $$f : U \to \mathbb{R}$$
>
> *The* ***derivative*** *of* f *at an* ***interior point*** $\underline{a} \in U$ *along the direction of* $\underline{y} \in \mathbb{R}^n$ *is:*
>
> $$f'(\underline{a};\underline{y}) = D_{\underline{y}} f(\underline{a}) = \lim_{t \to 0} \frac{f(\underline{a} + t\underline{y}) - f(\underline{a})}{t}$$
>
> *If* $\underline{y}$ *is a* ***unit vector***, $f'(\underline{a};\underline{y})$ *is a* ***directional derivative***.

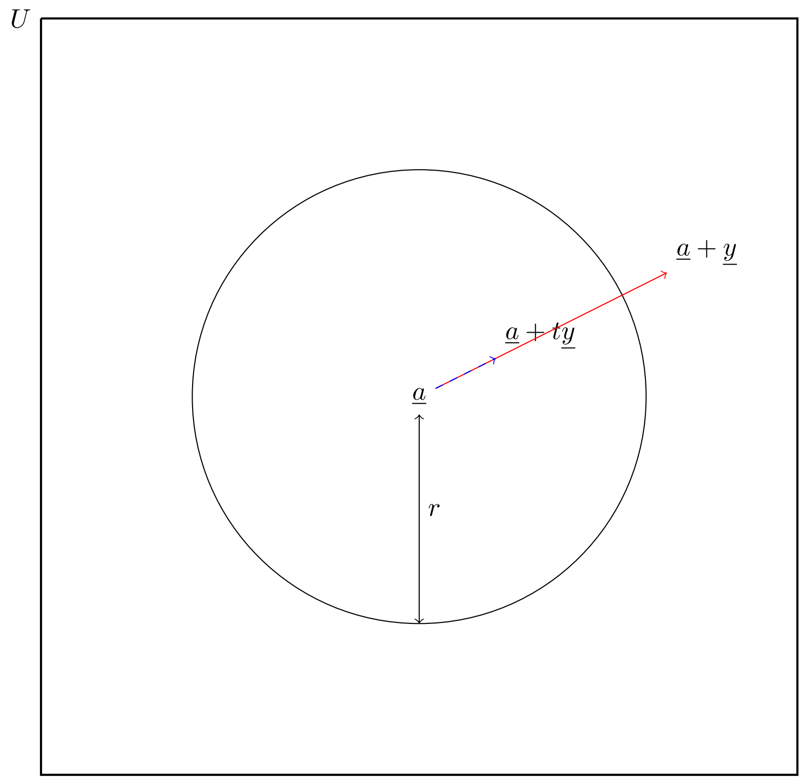


Figure 1: Operating in higher dimensions, derivatives depend on the **direction** which we pick, since the function changes differently in different directions. We can ensure that $\underline{a} + t\underline{y}$ remains within the ball $B_r(\underline{a})$ by enforcing $0 \leq t < \frac{r}{\|\underline{y}\|}$.

- **What is a partial derivative?**
  - a **directional derivative**, whereby we differentiate in the direction of the **canonical basis vectors** of $\mathbb{R}^n$
  - if we use
  $$x^1, x^2, \ldots, x^i, \ldots, x^n$$
  as the coordinate axes, with corresponding (canonical) basis vectors:
  $$\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_i, \ldots, \underline{e}_n$$
  we define:
  $$f'(\underline{a}; \underline{e}_i) \equiv \frac{\partial f}{\partial x^i}(\underline{a})$$
- **What is the gradient vector?**
  - a vector ∇f, where the ith component is the partial derivative $\frac{\partial f}{\partial x^i}$
  - technically, the components of ∇f are themselves functions
- **How do we compute directional derivatives using the gradient vector?**
  - given a vector $\underline{y}$, the directional derivative at $\underline{a}$ in the direction of $\underline{y}$ can be computed via:
  $$f'(\underline{a}; \underline{y}) = D_{\underline{y}} f(\underline{a}) = (\nabla f \cdot \underline{y})(\underline{a}) = \left\langle \nabla f, \underline{y} \right\rangle (\underline{a})$$

2.2 Definition: Differentiability of a Scalar Field

*A scalar field $f : U \to \mathbb{R}$ is **differentiable** at an interior point $\underline{a} \in U$ if its **total derivative** exists at $\underline{a}$. [Definition A.2]*

*The **total derivative** of f at $\underline{a}$ is a **unique** linear map:*

$$Df(\underline{a}) : \mathbb{R}^n \to \mathbb{R}$$

such that:

$$f(\underline{a} + \underline{v}) = f(\underline{a}) + Df(\underline{a})(\underline{v}) + \|\underline{v}\|E(\underline{a}, \underline{v}), \qquad \forall \underline{v} \in \mathbb{R}^n$$

*Here, $E(\underline{a}, \underline{v})$ is an **error term**, such that:*

$$\lim_{\|\underline{v}\| \to 0} E(\underline{a}, \underline{v}) = 0$$

- **How does the total derivative relate to the directional derivatives of a scalar field?**
  - assuming the **total derivative** exists for f, then:
  $$f(\underline{a} + \underline{v}) = f(\underline{a}) + Df(\underline{a})(\underline{v}) + \|\underline{v}\|E(\underline{a}, \underline{v}), \qquad \forall \underline{v} \in \mathbb{R}^n$$
  - if we rewrite $\underline{v} = t\underline{y}$, for some free parameter t:
  $$f(\underline{a} + t\underline{y}) = f(\underline{a}) + Df(\underline{a})(t\underline{y}) + \|t\underline{y}\|E(\underline{a}, t\underline{y})$$
  - if we exploit the linearity of the norm and total derivative:
  $$f(\underline{a} + t\underline{y}) = f(\underline{a}) + tDf(\underline{a})(\underline{y}) + |t|\|\underline{y}\|E(\underline{a}, t\underline{y})$$
  - but now if we rearrange the expression, and divide through by t:
  $$\frac{f(\underline{a} + t\underline{y}) - f(\underline{a})}{t} = Df(\underline{a})(\underline{y}) + \frac{|t|}{t}\|\underline{y}\|E(\underline{a}, t\underline{y})$$
  - taking the limit as $t \to 0$, and noticing that as $t \to 0$, then $\|t\underline{y}\| = \|\underline{v}\| \to 0$:
  $$\lim_{t \to 0} \frac{f(\underline{a} + t\underline{y}) - f(\underline{a})}{t} = f'(\underline{a}; \underline{y}) = Df(\underline{a})(\underline{y})$$
  - in other words, if the total derivative exists, then so do all other directional derivatives (since the total derivative evaluated at $\underline{y}$ is precisely the directional derivative of f in the direction of $\underline{y}$)
- **What does the directional derivative tell us about directional derivatives as function approximators?**
  - the total derivative is the best linear approximator for a scalar field close to some point $\underline{a}$
  - but since the directional derivatives are nothing but the result of evalutating the total derivative $Df(\underline{a})(\underline{y})$, this implies that in fact the directional derivatives provide the best linear approximation of f close to $\underline{a}$
  - this corresponds to the notion that directional derivatives span a hyperplane at $\underline{a}$ which best approximates the behaviour of f at said point

2.3 Remark: Justifying Differentiability

> *In 1 dimension, a function is* ***differentiable*** *at a point if its derivative exists at said point.*
> *However, when dealing with several dimensions, the "natural" extension to this won't work: we can't say a scalar field is* ***differentiable*** *at a point if* ***all*** *its partial derivatives are defined at said point.*
> *This is because we know that "****if a function is differentiable at a point, then it is continuous****". However, it is possible to construct a scalar field, with* ***all*** *directional derivatives defined at a point, but which isn't* ***continuous*** *at said point, which shows it won't be* ***differentiable****.*
> *More on this can be seen in this Quora article, and an example is provided below.*

2.4 Example: Partial Derivatives Defined, but not Continuous

Consider the scalar field:

$$f(x,y) = \begin{cases} \frac{xy^2}{x^2+y^4}, & \underline{x} \neq \underline{0} \\ 0, & \underline{x} = \underline{0} \end{cases}$$

The gradient vector (when $\underline{x} \neq \underline{0}$ is:

$$\nabla f = \begin{pmatrix} \frac{y^2(-x^2+y^4)^2}{(x^2+y^4)^2} \\ \frac{2xy(x^2-y^4)^2}{(x^2+y^4)^2} \end{pmatrix}$$

So the directional derivative at $\underline{a}$ in the direction of $\underline{u} = (u_1, u_2)^T$ is:

$$f'(x,y;\underline{u}) = \frac{y^2(-x^2+y^4)^2}{(x^2+y^4)^2}u_1 + \frac{2xy(x^2-y^4)^2}{(x^2+y^4)^2}u_2$$

If $\underline{x} = \underline{0}$, using the definition of directional derivative:

$$\begin{aligned} f'(\underline{0};\underline{u}) &= \lim_{t\to 0} \frac{f(\underline{0}+t\underline{u}) - f(\underline{0})}{t} \\ &= \lim_{t\to 0} \frac{f(t\underline{u})}{t} \\ &= \lim_{t\to 0} \frac{\frac{t^3u_1u_2^2}{t^2u_1^2+t^4u_2^4}}{t} \\ &= \lim_{t\to 0} \frac{tu_1u_2^2}{tu_1^2+t^3u_2^4} \\ &= \lim_{t\to 0} \frac{u_1u_2^2}{u_1^2+t^2u_2^4} \end{aligned}$$

Now, if $u_1 \neq 0$, then:

$$f'(\underline{0};\underline{u}) = \lim_{t\to 0} \frac{u_1u_2^2}{u_1^2+t^2u_2^4} = \frac{u_2^2}{u_1}$$

If $u_1 = 0$, then we have an indeterminate form, but L'Hopital's Rule tells us that the limit is 0.

Thus, we can see that f has well defined directional derivatives for any vector $\underline{u}$.

However, f isn't continuous at the origin. To show this, it is sufficient to show that the limit obtained by moving along 2 different curves which pass through the origin is different. Indeed, if we approach the origin via straight lines, f seems to be continuous:

$$f(x, mx) = \frac{m^2x^3}{x^2 + m^4x^4} = \frac{m^2x}{1 + m^4x^2}$$

So taking the limit as $(x, y) \to (0, 0)$:

$$\lim_{(x,y)\to(0,0)} \frac{m^2x}{1 + m^4x^2} = 0$$

so we get that the limit is 0 along any straight line trhough the origin, and this is the value of f at the origin, so f is continuous along these paths.

However, if we use a parabolic path $x = my^2$:

$$f(my^2, y) = \frac{my^4}{m^2y^4 + y^4} = \frac{m}{m+1} \neq 0$$

Hence, since the limits differ, f isn't continuous at the origin, even though all its partial derivatives are defined there.

> *This goes to show how the requirement of all partial derivatives being defined is not **sufficient** - as we will see below, we in fact require that the partial derivatives be defined **in the neighbourhood** of the point; that is, they must also be **continuous**.*

2.5 Theorem: Differentiability Implies Continuity

> *if $f : U \to \mathbb{R}$ is **differentiable** (in the sense that its total derivative exists) at an **interior point** $\underline{a} \in U$, then f is **continuous** at $\underline{a}$. (Theorem A.3)*

Proof. Since f is differentiable at $\underline{a}$:

$$f(\underline{a} + \underline{v}) = f(\underline{a}) + Df(\underline{a})(\underline{v}) + \|\underline{v}\|E(\underline{a}, \underline{v}), \qquad \forall \underline{v} \in \mathbb{R}^n$$

If we rearrange:

$$f(\underline{a} + \underline{v}) - f(\underline{a}) = \langle \nabla f, \underline{v} \rangle + \|\underline{v}\|E(\underline{a}, \underline{v}), \qquad \forall \underline{v} \in \mathbb{R}^n$$

Taking the absolute value of both sides:

$$|f(\underline{a} + \underline{v}) - f(\underline{a})| = |\langle \nabla f, \underline{v} \rangle + \|\underline{v}\|E(\underline{a}, \underline{v})|, \qquad \forall \underline{v} \in \mathbb{R}^n$$

If we then apply the Cauchy-Schwarz Inequality, followed by the Triangle Inequality on the RHS:

$$0 \leq |f(\underline{a} + \underline{v}) - f(\underline{a})| \leq \|\nabla f\|\|\underline{v}\| + \|\underline{v}\||E(\underline{a}, \underline{v})|, \qquad \forall \underline{v} \in \mathbb{R}^n$$

Then, as $\|v\| \to 0$, also $\|\nabla f\|\|\underline{v}\| + \|\underline{v}\||E(\underline{a}, \underline{v})|| \to 0$, so by Squeeze Theorem:

$$|f(\underline{a} + \underline{v}) - f(\underline{a})| \to 0 \implies f(\underline{a} + \underline{v}) \to f(\underline{a})$$

so f is continuous at $\underline{a}$, as required.

□

2.6 Continuous Differentiability

- **When is a function continuously differentiable at a point?**
  - let $f : U \to \mathbb{R}$ and consider $\underline{a} \in U$
  - if:
    * f is **differentiable** at $\underline{a}$
    * all the partial derivatives of f at $\underline{a}$ are continuous

    then f is **continuously differentiable** at $\underline{a}$
- **When is a function differentiable?**
  - when it is differentiable at every point in its domain
- **When is a function continuously differentiable?**
  - when it is continuously differentiable at every point in its domain
  - then, we say that $f : U \to \mathbb{R}$ is such that $f \in C^1(U)$, the class of continuously differentiable function on U

2.7 Theorem: Sufficient Condition for Differentiability

Let $f : U \to \mathbb{R}$, and let $\underline{a} \in U$ be an interior point. If:

1. *all **partial derivatives** exist at $\underline{a}$*
2. *all **partial derivatives** are **continuous** at $\underline{a}$*

*then f is **differentiable** at $\underline{a}$.*
(Theorem A.4)

2.8 Differentiability and Vector-Valued Functions

2.8.1 Vector-Valued Functions

- **What is a vector-valued function?**
  - a mapping:
  $$f : U \to \mathbb{R}^m$$
  where $U \subset \mathbb{R}^n$, and n, m need not be the same
  - we can think of vector-valued functions as **vectors** containing scalar fields as components:
  $$f = (f^1, \dots, f^m) = \sum_{i=1}^{m} f^i \underline{e}_i$$
  where:
  $$f^i : U \to \mathbb{R}$$
- **How do we defined the derivative of a vector valued function?**
  - the derivative is itself a vector, obtained by **componentwise** differentiation
  - formally, at an interior point $\underline{a} \in U$, the derivative in the direction of $\underline{y} \in \mathbb{R}^n$ is:
  $$f'(\underline{a}; \underline{y}) = \lim_{t \to 0} \frac{f(\underline{a} + t\underline{y}) - f(\underline{a})}{t} = (f^{1'}(\underline{a}; \underline{y}), f^{2'}(\underline{a}; \underline{y}), \dots, f^{m'}(\underline{a}; \underline{y}))^T$$

2.8.2 Definition: Differentiability of a Vector-Valued Function

Let $f : U \to \mathbb{R}^m, U \subset \mathbb{R}^n$.
*f is **differentiable** at an **interior point** $a \in U$ if its **total derivative** exists.*
*That is, there exists a **linear map** $Df(\underline{a}$ such that:*

$$f(\underline{a}+\underline{v}) = f(\underline{a}) + Df(\underline{a})(\underline{v}) + \|\underline{v}\|E(\underline{a},\underline{v})$$

*Here, $E(\underline{a},\underline{v})$ is a **vector-valued error term** such that:*

$$\lim_{\underline{v}\to\underline{0}} E(\underline{a},\underline{v}) = \underline{0}$$

2.8.3 The Total Derivative and the Jacobian Matrix

- **How are the derivatives of a vector-valued function related to its total derivative?**
  - by defining $\underline{v} = t\underline{y}$ it can be shown that:

$$Df(\underline{a})(\underline{y}) = f'(\underline{a};\underline{y}) = (f^{1'}(\underline{a};\underline{y}), f^{2'}(\underline{a};\underline{y}), \dots, f^{m'}(\underline{a};\underline{y}))^T$$

  - that is, derivatives provide the best linear approximation for the vector-valued function
- **What is the Jacobian Matrix?**
  - a generalisation of the **gradient vector** for vector-valued function
  - the **matrix representation** of the total derivative $Df(\underline{a})$
  - with respect to the canonical basis of $\mathbb{R}^n$ we have:

$$\begin{aligned}
Df(\underline{a})(\underline{y}) &= (f^{1'}(\underline{a};\underline{y}), f^{2'}(\underline{a};\underline{y}), \dots, f^{m'}(\underline{a};\underline{y}))^T \\
&= (\langle \nabla f^1, \underline{y}\rangle(\underline{a}), \dots, \langle \nabla f^m, \underline{y}\rangle(\underline{a})) \\
&= \sum_{j=1}^{m} \langle \nabla f^i, \underline{y}\rangle(\underline{a})\underline{e}_j \\
&= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} y^i \frac{\partial f^j}{\partial x^i}(\underline{a}) \right) \underline{e}_j \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} y^i \frac{\partial f^j}{\partial x^i}(\underline{a})\underline{e}_j
\end{aligned}$$

  - the $m \times n$ matrix $Df(\underline{a})$ is the matrix with entries defined by:

$$[Df(\underline{a})]^j_i = \frac{\partial f^j}{\partial x^i}(\underline{a})$$

  (here we consider the entry at row j and column i)

– alternatively, the Jacobian matrix is the matrix obtained by using the **gradient vector** of each f^i as rows:

$$Df(\underline{a}) = \begin{pmatrix} \nabla f^1(\underline{a})^T \\ \nabla f^2(\underline{a})^T \\ \vdots \\ \nabla f^m(\underline{a})^T \end{pmatrix} = \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(\underline{a}) & \frac{\partial f^1}{\partial x^2}(\underline{a}) & \dots & \frac{\partial f^1}{\partial x^n}(\underline{a}) \\ \frac{\partial f^2}{\partial x^1}(\underline{a}) & \frac{\partial f^2}{\partial x^2}(\underline{a}) & \dots & \frac{\partial f^2}{\partial x^n}(\underline{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f^m}{\partial x^1}(\underline{a}) & \frac{\partial f^m}{\partial x^2}(\underline{a}) & \dots & \frac{\partial f^m}{\partial x^n}(\underline{a}) \end{pmatrix}$$

2.8.4 Theorem: Sufficient Condition for Differentiability

Let $f : U \to \mathbb{R}^m, U \subset \mathbb{R}^n$, and let $\underline{a} \in \mathbb{R}^n$ be an interior point. If:

1. *the **Jacobian Matrix** exists at $\underline{a}$*
2. *all **partial derivatives** are **continuous** at $\underline{a}$*

then f is differentiable at $\underline{a}$.

2.8.5 Theorem: The Chain Rule

Consider:

$$f : U \to \mathbb{R}^m, \qquad U \subset \mathbb{R}^n$$

$$g : V \to \mathbb{R}^n, \qquad V \subset \mathbb{R}^p$$

$$h = f \circ g : W \to \mathbb{R}^m, \qquad W = V \cap g^{-1}(U) \subset \mathbb{R}^p$$

(we required $g^{-1}(U)$ to defined W to ensure that if $\underline{w} \in W$, then $f(g(\underline{w}))$ is defined, since $g(\underline{w}) \in U$ by construction).
Now, assume that:

- *$\underline{a}$ is an interior point of W*
- *g is **differentiable** at $\underline{a}$*
- *f is **differentiable** at $\underline{b} = g(\underline{a})$*

*Then, h is **differentiable** at $\underline{a}$, and the **derivative** is given by the **chain rule**:*

$$Dh(\underline{a}) = Df(\underline{b}) \circ Dg(\underline{a})$$

*Thinking of operators as matrices, we obtain the **Jacobian** of h by **matrix multiplication** of the jacobians of f, g:*

$$[Dh(\underline{a})]^j_k = \sum_{i=1}^{n} [Df(\underline{b})]^j_i [Dg(\underline{a})]^i_k$$

where:

- *$Dh(\underline{a})$ is a $m \times p$ matrix*
- *$Df(\underline{b})$ is a $m \times n$ matrix*
- *$Dg(\underline{a})$ is a $n \times p$ matrix*

(Theorem A.6)

2.9 Change of Coordinates: Worked Example

Thus far we have worked over the canonical basis; however, this need not be the basis of choice for certain problems. For instance, when working with circles/circular symmetry, polar coordinates might be more convenient:

$$x(r, \theta) = r\cos(\theta) \qquad y(r, \theta) = r\sin(\theta)$$

$$r(x, y) = \sqrt{x^2 + y^2} \qquad \theta(x, y) = \arctan\left(\frac{y}{x}\right)$$

We note:

- the radius is defined on the range $0 < r < \infty$
- the argument is defined on the range $\theta_0 \leq \theta < \theta_0 + 2\pi$, where θ_0 is any angle in radians

- this allows us to describe $\mathbb{R}^2 \setminus \{0\}$, since $(0,0)$ can't be described by a single argument

Now, if we want to operate over polar coordinates, but f uses x, y as arguments, how can we determine the derivative of f with respect to (r, θ)?

If $f : U \to \mathbb{R}, U \subset \mathbb{R}^2 \setminus \{0\}$, define:

$$g(r,\theta) = f(r\cos\theta, r\sin\theta)$$

We define g so that we have an explicit dependence on polar coordinates, and thus, can define partial derivatives.

Let $\underline{a}$ be a point described in polar coordinates, and $\underline{b}$ the corresponding point in cartesian coordinates. If we apply the chain rule, we know that:

$$\frac{\partial g}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r}$$
$$\frac{\partial g}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta}$$

We can compute:

$$\frac{\partial x}{\partial r} = \cos\theta$$
$$\frac{\partial x}{\partial \theta} = -r\sin\theta$$
$$\frac{\partial y}{\partial r} = \sin\theta$$
$$\frac{\partial t}{\partial \theta} = r\cos\theta$$

So:

$$\frac{\partial g}{\partial r}(\underline{a}) = \frac{\partial f}{\partial x}(\underline{b})\cos\theta + \frac{\partial f}{\partial y}(\underline{b})\sin\theta = \langle \nabla f(\underline{b}), \underline{e}_r\rangle = f'(\underline{b}; \underline{e}_r)$$
$$\frac{\partial g}{\partial \theta}(\underline{a}) = -r\frac{\partial f}{\partial x}(\underline{b})\sin\theta + r\frac{\partial f}{\partial y}(\underline{b})\cos\theta \implies \frac{1}{r}\frac{\partial g}{\partial \theta}(\underline{a}) = \langle \nabla f(\underline{b}), \underline{e}_\theta\rangle = f'(\underline{b}; \underline{e}_\theta)$$

where we use:

$$\underline{e}_r = \cos\theta \underline{e}_1 + \sin\theta \underline{e}_2$$
$$\underline{e}_\theta = -\sin\theta \underline{e}_1 + \cos\theta \underline{e}_2$$

In other words, if we use $\underline{e}_r, \underline{e}_\theta$ as basis vectors (we can verify they are orthonormal), this tells us that the partial derivatives of g are nothing but directional derivatives in another basis.

We can then use this to compute the gradient in polar coordinates. Notice, since e_r, e_θ defines a basis, we can write any vector $\underline{y}$ in terms of it:

$$\underline{y} = y^r \underline{e}_r + y^\theta \underline{e}_\theta$$

where:

$$y^\theta = \langle \underline{y}, e_r \rangle \qquad y^\theta = \langle \underline{y}, e_\theta \rangle$$

Now, if we consider the directional derivative:

$$\begin{aligned} f'(\underline{b}; \underline{y}) &= \langle \nabla f(\underline{b}), \underline{y} \rangle \\ &= \langle \nabla f(\underline{b}), y^r \underline{e}_r + y^\theta \underline{e}_\theta \rangle \\ &= y^r \langle \nabla f(\underline{b}), \underline{e}_r \rangle + y^\theta \langle \nabla f(\underline{b}), \underline{e}_\theta \rangle \\ &= y^r \frac{\partial g}{\partial r}(\underline{a}) + y^\theta \frac{1}{r}\frac{\partial g}{\partial \theta}(\underline{a}) \\ &= \langle \nabla g(\underline{a}), \underline{y} \rangle \end{aligned}$$

So we can conclude that in polar coordinates (with respect to the new basis):

$$\nabla g = \frac{\partial g}{\partial r}\underline{e}_r + \frac{1}{r}\frac{\partial g}{\partial \theta}\underline{e}_\theta$$

3 Level Sets and Implicitly Defined Functions

3.1 Level Sets and Regular Points

- **What is a level set?**
  - consider a scalar field $f : U \to \mathbb{R}, U \subset \mathbb{R}^n$
  - for some $c \in \mathbb{R}$ in the co-domain of f, define the level set $L(c)$ as:

$$L(c) = \{\underline{x} \mid \underline{x} \in U,\ f(\underline{x}) = c\}$$

  - for example, for $f(x, y, z) = x^2 + y^2 + z^2$, $L(1)$ is precisely the surface of a sphere in 3 dimensions of radius 1

- **What is a regular point?**
  - a point $\underline{a} \in L(c)$ such that:

$$\nabla f(\underline{a}) \neq \underline{0}$$

- **What do level sets represent?**
  - they represent a **hypersurface** in space
  - also known as **codimension** 1 surfaces:
    * the **codimension** corresponds to the number of **linearly independent** normal vectors of the surface
    * with one constraint $f(\underline{x}) = c$, we defined a surface in $\mathbb{R}^n$, which only has 1 normal vector, and so, is **codimension 1**
    * with 2 constraints $f(\underline{x}) = c, g(\underline{x}) = k$, we embed a lower dimensional surface (from $\mathbb{R}^{n-1}$) into $\mathbb{R}^n$, so it has to lid normal vectors, and so, is **codimension 2**
    * for example, in $\mathbb{R}^3$ the surface obtained by the intersection of 2 surfaces derived from constraints gives a curve from $\mathbb{R}^2$, but embedded in $\mathbb{R}^3$; such a curve has 2 normal vectors, which will be perpendicular
  - it is possible that if $f(x^1, \dots, x^n) = c$, we can sometimes solve to express a variable in terms of the other variables:

$$x^n = g(x^1, \dots, x^{n-1})$$

    Then, we can think of the surface as the **graph** of f
  - however, this representation is only **local**: with $f(x, y, z) = x^2 + y^2 + z^2$, we only get a local representation of the sphere:
    * if $z > 0$, $z = \sqrt{R^2 - x^2 - y^2}$ is the **northern hemisphere**
    * if $z < 0$, $z = -\sqrt{R^2 - x^2 - y^2}$ is the **southern hemisphere**
    * however, it is impossible to define the **whole** sphere through this representation
    * here, we have defined $z(x, y)$ **implicitly** by $f(x, y, z) = R^2$

- **How can we determine the tangent plane to a level set at a point?**
  - consider any $\underline{a} \in L(c)$
  - since the tangent to $L(c)$ at $\underline{a}$ is orthogonal to ∇f, any $\underline{x} \in \mathbb{R}^n$ satisfying:

$$\langle (\underline{x} - \underline{a}), \nabla f \rangle = 0$$

    will be part of the tangent plane at $\underline{a}$

- **What is a regular surface?**
  - a level set in which the gradient is never 0

3.2 Exercise: Gradient at a Regular Point is Perpendicular to Tangent

Define a curve:

$$\underline{\gamma} : I \to U, \qquad I \subset \mathbb{R}$$

where γ is continuously differentiable, and parametrised by $t \in I$. $\gamma(t)$ lies entirely in the level set $L(c)$ - that is, $\forall t \in I$:

$$f(\underline{\gamma}(t)) = c$$

If $\underline{a}$ is a regular point, and $\underline{a} = \gamma(t_0)$, show that $\nabla f(\underline{a})$ is orthogonal to the tangent curve $\underline{\gamma}$ at t_0.

Since $f(\underline{\gamma}(t)) = c$, we can differentiate with respect to t, applying the chain rule:

$$\sum_{i=1}^{n} \frac{\partial f}{\partial \gamma^i} \frac{d\gamma^i}{dt} = 0 \implies \langle \nabla f, \underline{\gamma}' \rangle = 0$$

as required.

3.3 Theorem: Implicit Function Theorem in $\mathbb{R}^2$

The implicit function theorem gives us conditions under which we can solve for variables in implicitly defined functions, in terms of the other variables.

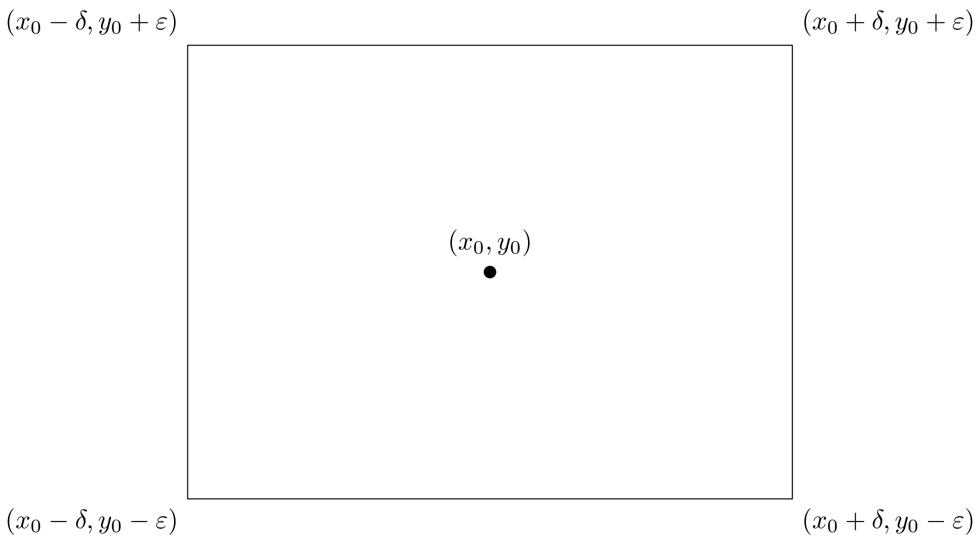
> *Let $f : U \to \mathbb{R}, U \subset \mathbb{R}^2$ be a* ***continuously differentiable*** *function. Define (x_0, y_0) to be an* ***interior point*** *of U, satisfying $f(x_0, y_0) = 0$. If $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$, then there exists a* ***neighbourhood*** *of (x_0, y_0) such that for x sufficiently close to x_0, there is a* ***unique continuously differentiable function*** *$y = F(x)$ such that:*
>
> - $y_0 = F(x_0)$
> - $f(x, F(x)) = 0$
> - $$F'(x) = -\frac{\frac{\partial f}{\partial x}(x, F(x))}{\frac{\partial f}{\partial y}(x, F(x))}$$
>
> *(Theorem A.10)*

Proof. Without loss of generality, since $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$, we can assume that $\frac{\partial f}{\partial y}(x_0, y_0) > 0$ (otherwise just repeat argument with $-f$).

Now, f is continuously differentiable, so all its partial derivatives are continuous, which means that $\frac{\partial f}{\partial y}(x, y) > 0$ for any (x, y) in a **neighbourhood** of (x_0, y_0).

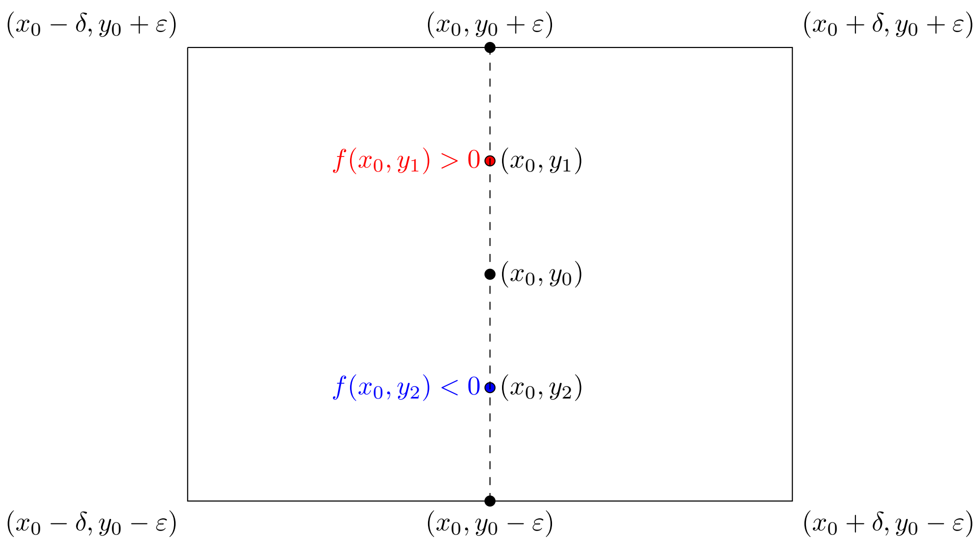
We can consider a rectangle within this neighbourhood, defined for $\delta, \varepsilon > 0$:

- $x_0 - \delta < x < x_0 + \delta$
- $y_0 - \varepsilon < y < y_0 + \delta$



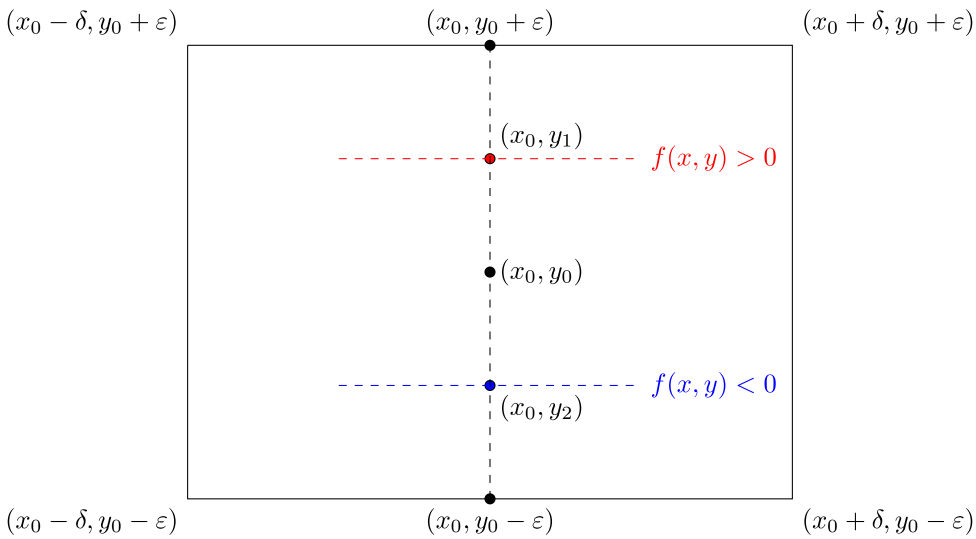
We can now consider varying y, whilst keeping $x = x_0$ fixed. In particular, since in the rectangle $\frac{\partial f}{\partial y}(x, y) > 0$, on the line $y_0 - \varepsilon < y < y_0 + \varepsilon$ $f(x_0, y)$ is an increasing function; since it attains a 0 at (x_0, y_0) it follows that:

$$\exists y_1 > y_0, y_2 < y_0 \in (y_0 - \varepsilon, y_0 + \varepsilon) \ : \ f(x_0, y_1) > 0, \quad f(x_0, y_2) < 0$$



But again by continuity of the partial derivative, in some interval around x_0, if we fix y_1, y_2 we will have that for $x \in (x_0 - \delta, x_0 + \delta)$:

- $f(x, y_1) > 0$
- $f(x, y_2) < 0$



In this neighbourhood, we still have $\frac{\partial f}{\partial y}(x, y) > 0$, so f will be strictly increasing; hence, for each x in the neighbourhood, we can always find a unique $y \in (y_2, y_1)$ such that $f(x, y) = 0$ (since we have a continuous, increasing function, which goes from negative to positive values).

Thus, in this neighbourhood, we have implicitly defined y as a function of x (since for each x, there is a unique y satisfying $f(x, y) = 0$), so we can write $y = F(x)$. By construction, we will have $y_0 = F(x_0)$.

Finally, by definition we have that $f(x, F(x)) = 0$ in the neighbourhood (this is the definition of F), so if we differentiate with respect to x:

$$\frac{\partial f}{\partial x}(x, F(x)) + \frac{\partial f}{\partial y}(x, F(x))F'(x) = 0 \implies F'(x) = -\frac{\frac{\partial f}{\partial x}(x, F(x))}{\frac{\partial f}{\partial y}(x, F(x))}$$

Notice, this doesn't outline the neighbourhood on which this relationship can be made explicit, nor does it guarantee that we can actually obtain the explicit expression $y = F(x)$. However, it does give us an explicit expression for the derivative of an implicitly defined function - what is known as **implicit differentiation**.

□

3.4 Theorem: The Implicit Function Theorem

4 The Hessian Matrix and Stationary Points

4.1 Critical Points

- **What is a critical point?**
  - let $f : U \to \mathbb{R}$ be a scalar field
  - an interior point $\underline{x}_0$ is a **critical point** (or **extremum**) if the directional derivatives of f vanish in all directions at $\underline{x}_0$
  - this implies that $f(\underline{x}_0)$ is a local maximum/minimum or a saddle point
- **How does a scalar field behave around a critical point?**

- consider a ball B around a critical point $\underline{x}_0$:
  * if $\forall \underline{x} \in B$, $f(\underline{x}) - f(\underline{x}_0) > 0$, then $\underline{x}_0$ is a **local minimum**
  * if $\forall \underline{x} \in B$, $f(\underline{x}) - f(\underline{x}_0) < 0$, then $\underline{x}_0$ is a **local maximum**
  * if the sign of $f(\underline{x}) - f(\underline{x}_0)$ depends on the direction in which we move away from $\underline{x}_0$, then $\underline{x}_0$ is a **saddle point**

4.2 The Hessian Matrix

- **What is the Hessian matrix?**
  - if f has **well-defined** and **continuous** second order partial derivatives at a point $\underline{a}$, we can expand:
  $$f(\underline{a} + \underline{v}) = f(\underline{a}) + Df(\underline{a})(\underline{v}) + \frac{1}{2}\underline{v}^T Hf(\underline{a})(\underline{v})\underline{v} + \|\underline{v}\|^2 E(\underline{a}, \underline{v})$$
  (this is similar to a second order Taylor expansion)
  - $E(\underline{a}, \underline{v})$ is an **error-term** which vanishes as $\underline{v} \to \underline{0}$
  - $Hf(\underline{a})$ is the **Hessian** of f at $\underline{a}$, and has entries:
  $$[Hf(\underline{a})]_{ij} = \frac{\partial^2 f}{\partial x^i \partial x^j}(\underline{a})$$
- **What are the properties of the Hessian matrix?**
  - if the second order partial derivatives are **continuous** at $\underline{a}$, then the **Hessian** is **symmetric**
  - this means that:
    * it has **real** eigenvalues
    * eigenvectors of different eigenvalues are **orthogonal**
    * it can be **diagonalised**:
    $$Hf(\underline{a}) = P^T DP$$
    where P is the matrix with eigenvectors as columns, and D is the diagonal matrix with eigenvalues in the diagonal
- **Why is the Hessian matrix useful for determining the nature of extrema?**
  - by definition:
  $$Df(\underline{a})(\underline{v}) = \langle \nabla f(\underline{a}), \underline{v} \rangle$$
  so at a critical point $\underline{x}_0$, we have that $Df(\underline{x}_0)(\underline{v}) = 0$
  - hence:
  $$f(\underline{x}_0 + \underline{v}) - f(\underline{x}_0) = \frac{1}{2}\underline{v}^T Hf(\underline{a})(\underline{x}_0)\underline{v} + \|\underline{v}\|^2 E(\underline{x}_0, \underline{v})$$
  - as $\underline{v} \to \underline{0}$, we can see that the sign of $f(\underline{x}_0 + \underline{v}) - f(\underline{x}_0)$ will **solely** depend on the **quadratic form** $\underline{v}^T Hf(\underline{a})(\underline{x}_0)\underline{v}$

4.3 Theorem: Nature of Extrema from Hessian

> *Let $f : U \to \mathbb{R}$ be a **scalar field** whose second order partial derivatives at an **extremum** $\underline{x}_0$ exist and are **continuous** in a ball around $\underline{x}_0$.*
> *The **extremum** at $\underline{x}_0$ is:*
>
> - *a **local minimum** if **all** the **eigenvalues** of $Hf(\underline{x}_0)$ are **positive***
> - *a **local maximum** if **all** the **eigenvalues** of $Hf(\underline{x}_0)$ are **negative***
> - *a **saddle point** if $Hf(\underline{x}_0)$ has both **positive and negative** eigenvalues*
> - ***inconclusive** if there is at least one zero eigenvalue, with all other eigenvalues of the **same** sign*

Proof. Diagonalising the Hessian, we can transform the quadratic form:

$$\underline{v}^T Hf(\underline{a})(\underline{x}_0)\underline{v} = \underline{v}^T(P^T DP)\underline{v} = \underline{w}^T D\underline{w}$$

where $\underline{w} = P\underline{v}$.

Since the diagonals of D are the eigenvalues of $Hf(\underline{a})(\underline{x}_0)$:

$$\underline{w}^T D\underline{w} = \sum_{i=1}^{n} \lambda_i (w^i)^2$$

Hence, it follows that since $(w^i)^2 \geq 0$:

- if all the eigenvalues are positive, $\underline{v}^T Hf(\underline{a})(\underline{x}_0)\underline{v} > 0$, and so $f(\underline{x}_0 + \underline{v}) - f(\underline{x}_0) > 0$ so x_0 is a local minimum
- if all the eigenvalues are negative, $\underline{v}^T Hf(\underline{a})(\underline{x}_0)\underline{v} < 0$, and so $f(\underline{x}_0 + \underline{v}) - f(\underline{x}_0) < 0$ so x_0 is a local maximum

□

5 The Method of Lagrange Multipliers

5.1 Lagrange Multipliers

- **What is a constrained optimisation problem?**
  - the problem of finding an **extremum** of a function, subject to **constraints**
  - typically, we seek to find an extremum of $f(\underline{x})$, given that the variables (domain) are constrained to satisfy $g(\underline{x}) = c$ (can be simplified to $g(\underline{x}) = 0$, by defining $g(\underline{x}) := g(\underline{x}) - c$)
- **How can one go about solving a constrained optimisation problem?**
  - given $g(\underline{x}) = 0$, we can solve the constraint for one variable:

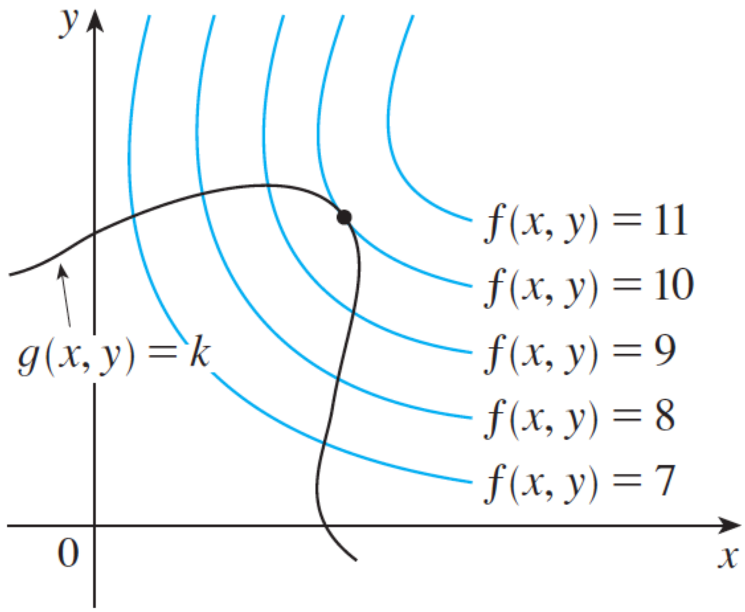
$$x^n = F(x^1, \ldots, x^{n-1})$$

- then, we can optimise:

$$f(x^1, \dots, x^{n-1}, F(x^1, \dots, x^{n-1}))$$

  by the previous methods (i.e Hessian)
- however, F will only apply locally, so this approach might not work in general

- **What is the method of Lagrange multipliers?**
  - method providing **necessary** conditions for a point to be a critical point



  - we seek an extremum for $f(\underline{x})$, given that $\underline{x}$ must lie (satisfy) within the curve $g(\underline{x}) = 0$
  - this is equivalent to finding the largest value of c such that the level curve $f(\underline{x}) = c$ intersects $g(\underline{x}) = 0$
  - intuitively, this occurs precisely when the 2 level curves intersect **tangentially**
  - but if the curves are tangential, this is equivalent to the **gradient vectors** being parallel. In other words, at an extremum $\underline{x}_0$, there exists $\lambda \in \mathbb{R}$ such that:

$$\nabla f(\underline{x}) = \lambda \nabla g(\underline{x})$$

    (since both gradient vectors are perpendicular to the tangent)
  - with this, we can find all $\underline{x}_0$ satisfying the Lagrange multipliers, and then use the Hessian to verify the nature of the points

- **What is the Lagrangian?**
  - in practice, when applying the method of Lagrange multipliers, we typically use the Lagrangian:

$$\mathcal{L}(\underline{x}, \lambda) = f(\underline{x}) - \lambda g(\underline{x})$$

  - if we compute the gradient:

$$\nabla \mathcal{L}(\underline{x}, \lambda) = \begin{pmatrix} \nabla f(\underline{x}) - \lambda \nabla g(\underline{x}) \\ ----- \\ g(\underline{x}) \end{pmatrix}$$

- notice, at a critical point $\underline{x}_0$ which satisfies the constraint, we get:

$$\nabla \mathcal{L}(\underline{x}_0, \lambda) = \underline{0}$$

- in other words, any critical point of λ will be a critical point of f given the constraint g

5.2 Theorem: Method of Lagrange Multipliers

Let:

$$f, g : U \to \mathbb{R}, \qquad U \subset \mathbb{R}^n$$

Let $\underline{x}_0$ be an interior point of U, such that $\underline{x}_0$ is an ***extremum*** *of f, subject to $g(\underline{x}) = 0$.*
Define the set of all $\underline{x}$ satisfying the constraint as U_0:

$$U_0 = \{\underline{x} \mid \underline{x} \in U, g(\underline{x}) = 0\}$$

Assume there exists an n-ball $B(\underline{x}_0)$, such that:

- $f(\underline{x}) \leq f(\underline{x}_0), \quad \forall x \in U_0 \cap B(\underline{x}_0)$
- *or* $f(\underline{x}) \geq f(\underline{x}_0), \quad \forall x \in U_0 \cap B(\underline{x}_0)$

Then, if $\nabla g(\underline{x}_0) \neq 0$, $\exists \lambda in \mathbb{R}$ such that $\underline{x}_0$ is a critical point of the Lagrangian:

$$\mathcal{L} : U \times \mathbb{R} \to \mathbb{R}$$

$$\mathcal{L}(\underline{x}, \lambda) = f(\underline{x}) - \lambda g(\underline{x})$$

(Theorem A.12)

5.3 Lagrange Multipliers for Multiple Constraints

- **How do Lagrange Multipliers apply when there are multiple constraints?**
  - consider the scalar field:

$$f : \mathbb{R}^n \to \mathbb{R}$$

  - if there are m constraints (with $m < n$) we can encode them within a **vector-valued function**:

$$g : \mathbb{R}^n \to \mathbb{R}^m$$

  - then, we seek to find all $\underline{x} \in \mathbb{R}^n$ satisfying the m constrains:

$$g(\underline{x}) = \underline{0} \in \mathbb{R}^m$$

  - we can use a modified Lagrangian:

$$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$$

$$\mathcal{L}(\underline{x}, \lambda) = f(\underline{x}) - \langle \underline{\lambda}, g(\underline{x}) \rangle, \qquad \underline{\lambda} \in \mathbb{R}^m$$

    to determine all possible critical points
  - if $(\underline{x}_0, \underline{\lambda}_0)$ is a critical point of $\mathcal{L}$ then:

    * $Df(\underline{x}_0) = \langle \lambda_0, Dg(\underline{x}_0) \rangle$
    * $g(\underline{x}_0 = 0$
  - here, recall that $Dg(\underline{x}_0)$ is a **Jacobian Matrix**, with ∇g_i^T as row vectors; for this to yield an answer, we require that the matrix have rank m - in other words, that the gradients ∇g_i are LiD
  - also notice that the $\underline{x}_0$ can satisfy all the conditions, but not be a critical point of f - it can just be a linear combination of the gradients of g
  - less abstractly, if we have 2 constrains g_1, g_2 then we need to satisfy:
    * $\nabla f(\underline{x}) + \lambda_1 \nabla g_1(\underline{x}) + \lambda_2 \nabla g_2(\underline{x}) = \underline{0}$
    * $g_1(\underline{x}) = 0$
    * $g_2(\underline{x}) = 0$
  - more can be found in this article by the University of Toronto

5.4 Exercise: Applying Lagrange Multipliers

Find the maxima of the function $f(x, y) = xy$ subject to the constraint $x^2 + y^2 = 1$.

We begin by computing $\nabla f, \nabla g$:

$$\frac{\partial f}{\partial x} = y \qquad \frac{\partial f}{\partial y} = x \qquad \frac{\partial g}{\partial x} = 2x \qquad \frac{\partial g}{\partial y} = 2y$$

Lagrange multipliers tell us that for (x, y) to be a critical point:

$$y = \lambda 2x \qquad x = \lambda 2y$$

Substituting values in:

$$y = \lambda 2x \implies y = 4\lambda^2 y \implies \lambda = \pm\frac{1}{2}$$

Hence, the Lagrange multiplier can only be $\pm\frac{1}{2}$. We now satisfy the constraint, by using the fact that $y = \pm x$:

$$x^2 + y^2 = 1 \implies 2x^2 = 1 \implies x = \frac{1}{\sqrt{2}}$$

Hence, there are 4 (possible) critical points:

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \qquad \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \qquad \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \qquad \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$$

By inspection, it can be seen that $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$ lead to maximising f.

6 Curve Parametrisation, Arc Length and Regular Surfaces

6.1 Curve Parametrisation

- **What is a parametrised curve/surface?**
  - a way of describing a curve/surface by using a **parameter space**
  - simpler than defining implicitly
  - for example, a **sphere**:

    $$x^2 + y^2 + z^2 = R^2$$

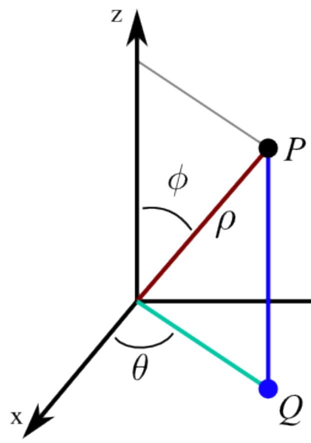
    can be parametrised using **spherical coordinates**:

    $$x(\phi, \theta) = R\cos\phi\sin\theta$$

$$y(\phi, \theta) = R \sin\phi \sin\theta$$

$$x(\phi, \theta) = R \cos\theta$$

where $0 \leq \phi < 2\pi$ and $0 < \theta < \pi$



6.2 The Arc-Length

- **How do we compute the arc length of a curve?**

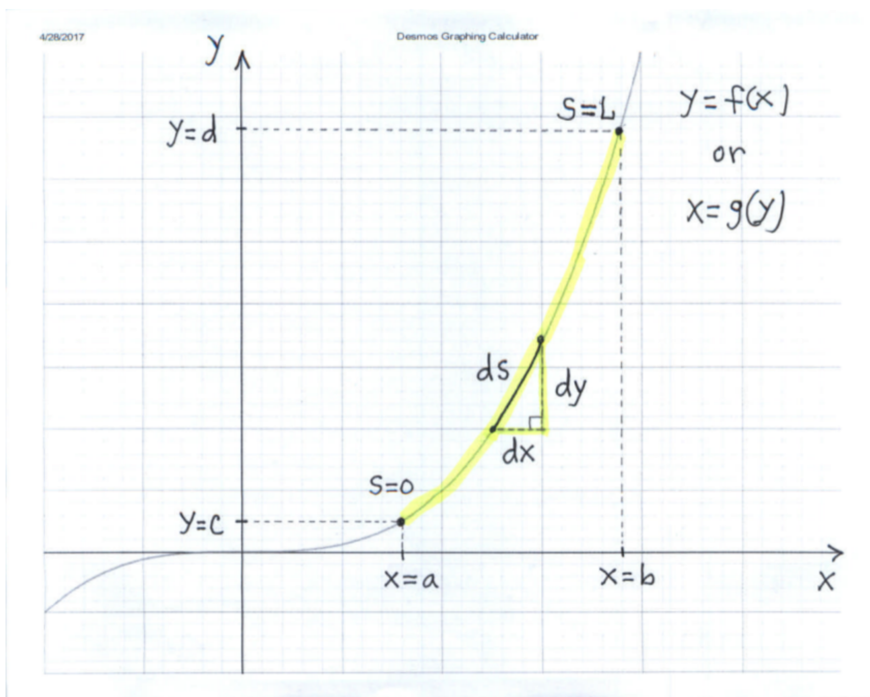


Figure 2: We can approximate each segment of the curve with a small segment $ds = dx^2 + dy^2$.

– for a explicit curve $y = f(x)$, the arc length between 2 points x_1, x_2:

$$L = \int_{x_1}^{x_2} \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx$$

– if the curve is parametrised, such that $x_1 = x(t_1)$ and $x_2 = x(t_2)$, we have:

$$L = \int_{x_1}^{x_2} \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx = \int_{t_1}^{t_2} \sqrt{1 + \left(\frac{dy/dt}{dx/dt}\right)^2} \frac{dx}{dt} dt = \int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt$$

– this can be written in much simpler form:

$$L = \int_{t_1}^{t_2} \|\dot{\underline{x}}(t)\| dt$$

– this assumes that the curve $\underline{x}(t)$ is **regular**: its derivative is non-zero $\forall t$

- **How does parametrisation affect the arc length?**
  – arc length is independent of parametrisation
  – for example, if we have the curve $y = x^2$, we would expect that:

$$(x(t), y(t)) = (t, t^2)$$
$$(x(t), y(t)) = (2t, 4t^2)$$

have the same arc length

  – because of this, wlg we could argue that the only interval that matters is $t \in [0, 1]$, since any other interval can be attained by reparametrising the curve

6.3 Exercise: Arc Length is Independent of Reparametrisation

Consider a parametrisation $\gamma(t), t \in [0, 1]$. Notice, since 2 parametrisations trace out the same curve, the only different between 2 parametrisations is how **quickly** they traverse the curve.

Hence, define $\tau(t)$ such that, for some other parametrisation $x(t)$:

$$\gamma(t) = x(\tau(t)), \qquad \forall t \in [0, 1]$$

Then:

$$\begin{aligned} \int_0^1 \|\gamma'(t)\| dt &= \int_0^1 \left\| \frac{d}{dt}(x(\tau(t)) \right\| dt \\ &= \int_0^1 \left\| x'(\tau) \frac{d\tau}{dt} \right\| dt \\ &= \int_{\tau(0)}^{\tau(1)} \|x'(\tau)\| d\tau \end{aligned}$$

so the parametrisation doesn't affect the value of the arc length.

6.4 Surface of Revolution

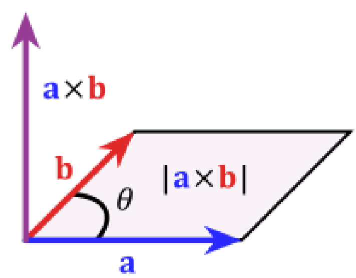
- **What is a surface of revolution?**
  – the surface obtained by rotating a curve around an axis
  – for a curve $(x, y(x))$, the surface area of such a surface for $x \in [x_0, x_1]$ is:

$$\int_{x_0}^{x_1} 2\pi y(x) \sqrt{1 + y'(x)^2} dx$$

  – intuitively, the surface area can be thought of as the sum of many small cylinders, where:
    * $y(x)$ gives the radius of the cylinder
    * $\sqrt{1 + y'(x)^2}$ gives the height of the cylinder

6.5 Surface Area

- **What is a regular surface in $\mathbb{R}^3$?**
  - a **continuously differentiable** map:
  $$x : U \subset \mathbb{R}^2 \to \mathbb{R}^3$$
  defined by:
  $$(u, v) \mapsto x(u, v)$$
  - for all $(u, v) \in U$, we require that the tangent vectors $\frac{\partial x}{\partial u}, \frac{\partial x}{\partial v}$ be **linearly independent**: they should **span** the **tangent plane** to the surface at any point (this is because the cross product of the vectors defines the tangent plane, and if they are lienarly dependent, the cross product will yield $\underline{0}$)
- **How do we compute the surface area of a regular surface?**
  - recall, the **norm** of a cross product gives the area of the parallelogram defined by the vectors in the cross product



  - from this, we derive the **surface area** of a regular surface:
  $$\iint \left\| \frac{\partial x}{\partial u} \times \frac{\partial x}{\partial v} \right\| du\ dv$$
  - alternatively:
  $$\iint \sqrt{\left\| \frac{\partial x}{\partial u} \right\|^2 \left\| \frac{\partial x}{\partial v} \right\|^2 - \left(\left\langle \frac{\partial x}{\partial u}, \frac{\partial x}{\partial v} \right\rangle \right)^2} du\ dv$$

If we use the parametrisation:

$$(u, v, z(u, v))$$

then the partial derivatives are:

$$\frac{\partial x}{\partial u} = \begin{pmatrix} 1 \\ 0 \\ \frac{\partial z}{\partial u} \end{pmatrix} \qquad \frac{\partial x}{\partial v} = \begin{pmatrix} 0 \\ 1 \\ \frac{\partial z}{\partial v} \end{pmatrix}$$

So the surface area becomes:

$$\iint \sqrt{\left\|\frac{\partial x}{\partial u}\right\|^2 \left\|\frac{\partial x}{\partial v}\right\|^2 - \left(\left\langle \frac{\partial x}{\partial u}, \frac{\partial x}{\partial v}\right\rangle\right)^2} du\, dv$$

$$= \iint \sqrt{\left(1 + \left(\frac{\partial z}{\partial u}\right)^2\right)\left(1 + \left(\frac{\partial z}{\partial v}\right)^2\right) - \left(\frac{\partial z}{\partial u}\right)^2 \left(\frac{\partial z}{\partial u}\right)^2} du\, dv$$

$$\iint \sqrt{1 + \left(\frac{\partial z}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2} du\, dv$$