§13.1-3: Extreme Values

In this week's notes, we develop techniques for finding (absolute) maxima and minima of multivariate functions, i.e. points $\mathbf{a}$ where $f(\mathbf{a}) \geq f(\mathbf{x})$ for all $\mathbf{x}$ in the domain of f (or $f(\mathbf{a}) \leq f(\mathbf{x})$, for minima). The value of f at the maxima and minima are called *extreme values*, or extrema. The function f that we're extremising is called the *objective function*.

As a start, imagine we wish to find the tallest person in the class. We can measure the heights of all 51 students and compare the measurements to find the tallest student.

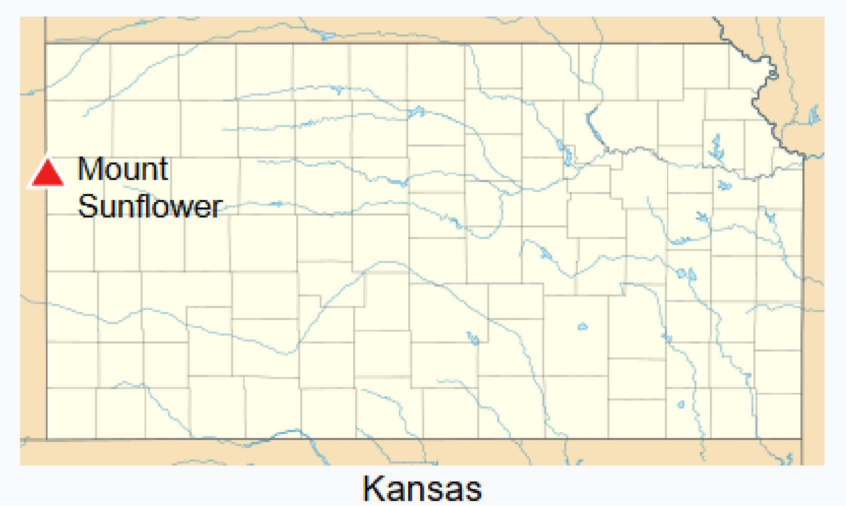
Now imagine we wish to find the highest point in Kansas (a state in USA). Now the domain of the height function is "Kansas", which contains infinitely many points, so it's impractical to measure the height at every point and compare. So instead, let's measure the height of all the mountain tops (all the local maxima) in Kansas and compare their heights.

But this is “Mount Sunflower”, the highest point in Kansas. It does not look like a mountain.

(pictures from Wikipedia and AtlasObscura)

But this is “Mount Sunflower”, the highest point in Kansas. It does not look like a mountain.

As Wikipedia explains, “the state of Kansas gradually increases in [height] from the east to the west”. The height continues to increase to the west of Mount Sunflower, but that area is in Colorado, not Kansas.



(pictures from Wikipedia and AtlasObscura)

Another way to explain this: let f be the height function. If $\nabla f(a,b) \neq \mathbf{0}$, then I can walk from (a,b) in the direction of $\nabla f(a,b)$ to increase my height. So, if (a,b) is the highest point in Kansas, one of three things must happen:

- $\nabla f(a,b) = \mathbf{0}$ so there is no direction I can walk in to climb higher;
- f is not differentiable at (a,b) (i.e. (a,b) is a *singular* point), so I don't know which direction to walk in to climb higher;
- (a,b) is on the boundary, and $\nabla f(a,b)$ "points out of the domain", so I must walk outside Kansas to climb higher.

Another way to explain this: let f be the height function. If $\nabla f(a,b) \neq \mathbf{0}$, then I can walk from (a,b) in the direction of $\nabla f(a,b)$ to increase my height. So, if (a,b) is the highest point in Kansas, one of three things must happen:

- $\nabla f(a,b) = \mathbf{0}$ so there is no direction I can walk in to climb higher;
- f is not differentiable at (a,b) (i.e. (a,b) is a *singular* point), so I don't know which direction to walk in to climb higher;
- (a,b) is on the boundary, and $\nabla f(a,b)$ "points out of the domain", so I must walk outside Kansas to climb higher.

This idea is true for functions of any number of variables:

Theorem 1: Necessary conditions for extreme values: If $f : \mathcal{D} \to \mathbb{R}$ achieves a maximum or minimum value at $\mathbf{a}$, then $\mathbf{a}$ must be a critical point of f, or a singular point of f, or a boundary point of D.

For single-variable functions, we handle the possibility of extreme values on the boundary by checking the endpoints of an interval (Homework 4 last question). For higher-dimensional domains, the boundaries are more complicated, so we cannot check every boundary point.

As stated in the theorem, if f achieves its extreme value at a non-boundary point $\mathbf{a}$, then either $\nabla f(\mathbf{a}) = \mathbf{0}$, or f is not differentiable at $\mathbf{a}$.

The main topic in this week's notes is to derive similar conditions that a boundary point must satisfy if f achieves its extreme value there. We will give two different set of conditions, which are useful for differently-shaped boundaries. Then we only need to check the shorter list of boundary points that satisfy these conditions.

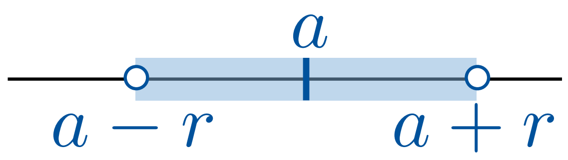
In this week's notes:

- What is the boundary? (p5-9, §10.1)
- General strategy for finding extreme values (p10, §13.1)
- Two ways to check for extreme values on the boundary:
  - Parametrisation: for straight lines and simple boundaries (p11-13, §13.2)
  - Lagrange multipliers: for complicated boundaries (p15-18, §13.3)
- Do extreme values exist? (p20-22, §13.1)

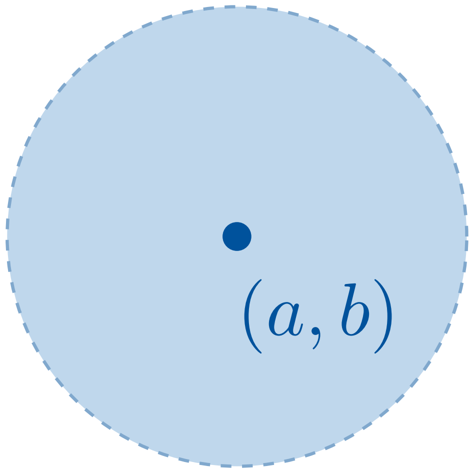
What is the boundary, and related ideas from topology (last page of §10.1)

Definition: Let P be a point in $\mathbb{R}^n$. The *ball* of radius r around P is the set of points whose distance from P is less than r.

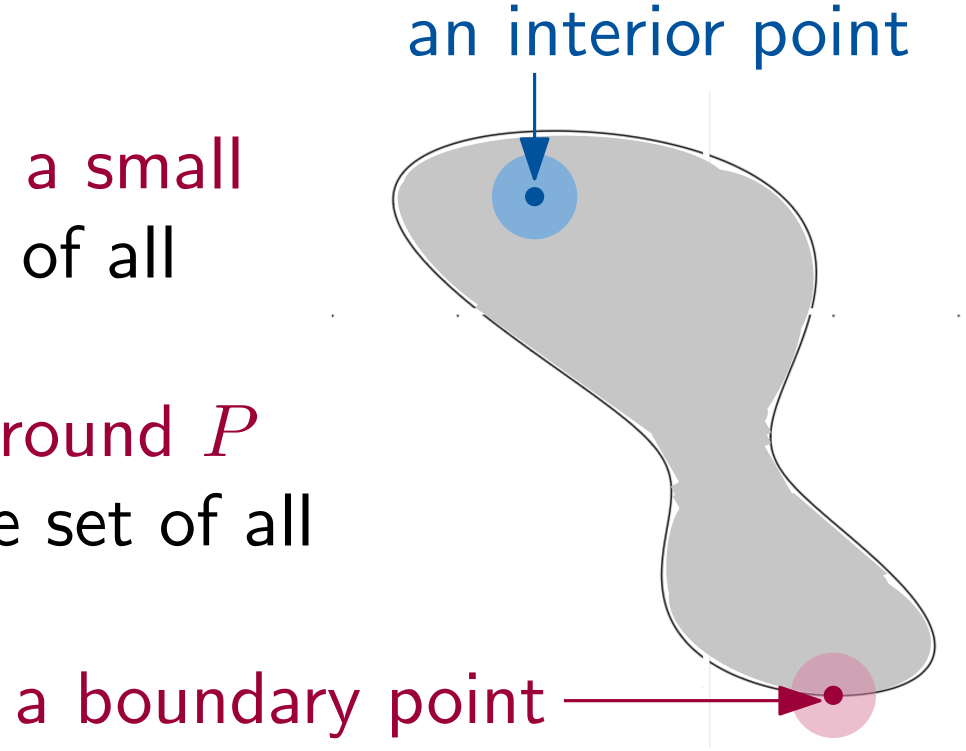
Example: In $\mathbb{R}^1$, the ball of radius r around a number a is the open interval $(a-r, a+r)$.



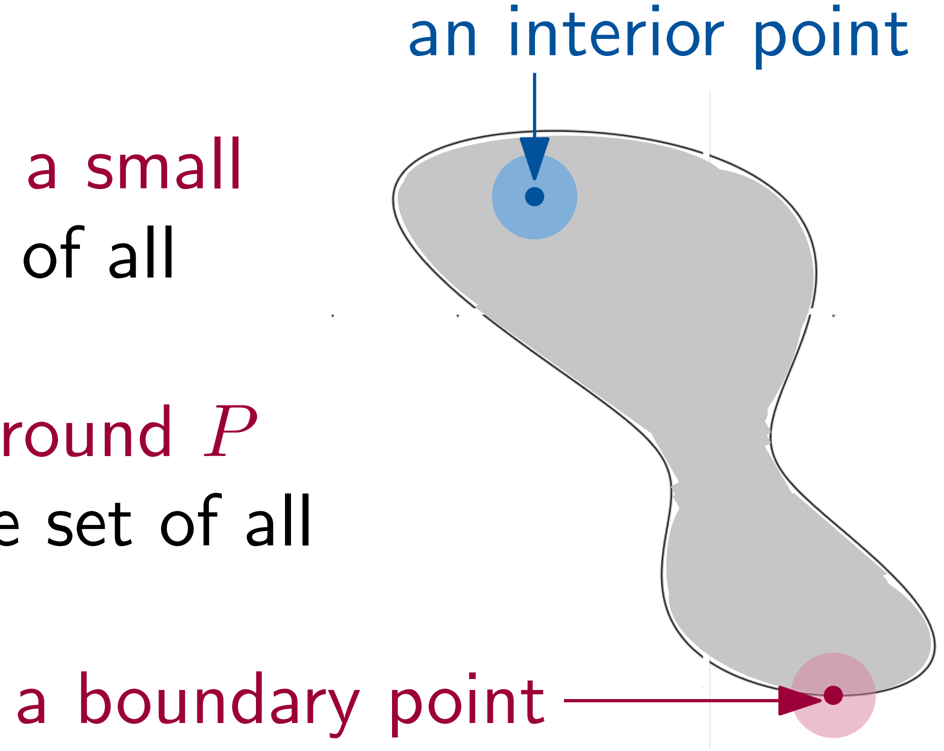
Example: In $\mathbb{R}^2$, the ball of radius r around a point (a,b) is the region inside the circle of radius r and centre (a,b): $\{(x,y) \in \mathbb{R}^2 \mid (x-a)^2 + (y-b)^2 < r^2\}$.



Definition: Let S be a subset of $\mathbb{R}^n$.
A point P in S is an *interior point* of S if there is a small ball around P that is completely inside S. The set of all interior points of S is the *interior of S*.
A point P is a *boundary point* of S if every ball around P contains both points in S and points not in S. The set of all boundary points of S is the *boundary of S*.



Definition: Let S be a subset of $\mathbb{R}^n$.
A point P in S is an *interior point* of S if there is a small ball around P that is completely inside S. The set of all interior points of S is the *interior of S*.
A point P is a *boundary point* of S if every ball around P contains both points in S and points not in S. The set of all boundary points of S is the *boundary of S*.

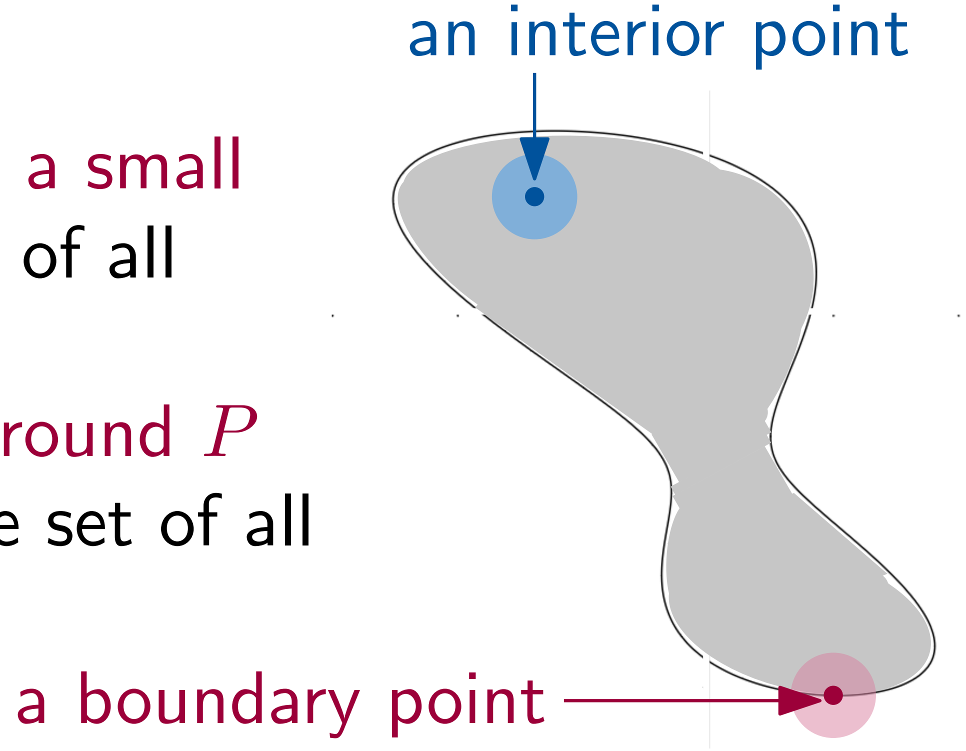


Example: Consider $S = (a, b]$.
The interior is (a, b).
The boundary are the two points $\{a\} \cup \{b\}$.
Note that a boundary point of S may be in S or not in S.

Definition: Let S be a subset of $\mathbb{R}^n$.

A point P in S is an *interior point* of S if there is a small ball around P that is completely inside S. The set of all interior points of S is the *interior of S*.

A point P is a *boundary point* of S if every ball around P contains both points in S and points not in S. The set of all boundary points of S is the *boundary of S*.

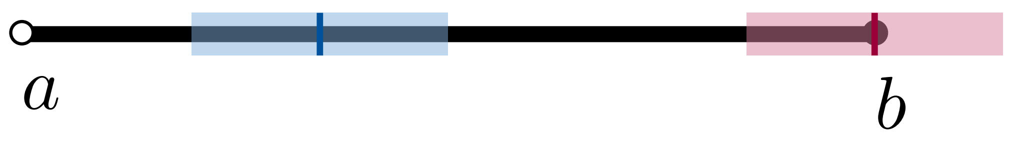


Example: Consider $S = (a, b]$.

The interior is (a, b).

The boundary are the two points $\{a\} \cup \{b\}$.

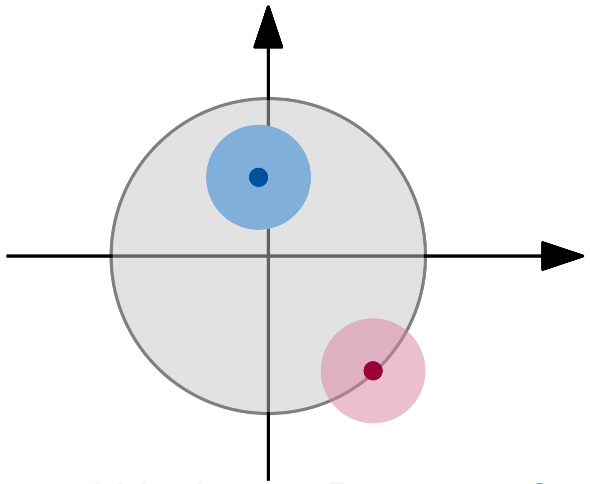
Note that a boundary point of S may be in S or not in S.



Example: Consider the unit disk $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$.

The interior is $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$.

The boundary is $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$.



As these examples show, if S is defined by inequalities, then its interior is usually the part with a $<$ or $>$ sign, and its boundary is the part with a $=$ sign.

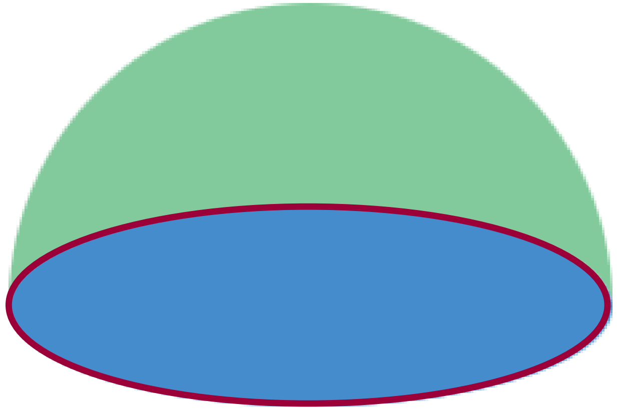
It is possible for a set to have no interior points e.g. the unit sphere $\left\{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2=1\right\}$, or any surface in $\mathbb{R}^3$ defined by one equation.

As these examples show, if S is defined by inequalities, then its interior is usually the part with a $<$ or $>$ sign, and its boundary is the part with a $=$ sign.

It is possible for a set to have no interior points e.g. the unit sphere $\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$, or any surface in $\mathbb{R}^3$ defined by one equation.

If S is defined by more than 1 inequality, then it is often convenient to write its boundary as the union of several sets ("boundary pieces") which are each defined by $<$, $>$ or $=$ signs (i.e. no $\leq$ or $\geq$ signs, because it is hard to search for extrema on this type of set).

Example: Consider the upper half ball $\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 \leq 1,\ z \geq 0\}$.
Its interior is $\{x^2 + y^2 + z^2 < 1,\ z > 0\}$.
Its boundary is in three pieces:
the "curved part on top": $\{x^2 + y^2 + z^2 = 1,\ z > 0\}$;
the "flat part on the bottom": $\{x^2 + y^2 + z^2 < 1,\ z = 0\}$;
the "circle on the edge": $\{x^2 + y^2 + z^2 = 1,\ z = 0\}$.



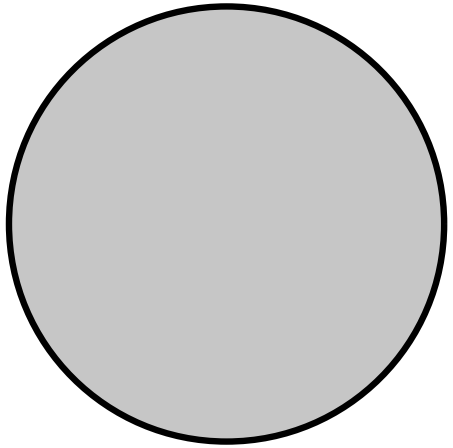
(picture from URegina MathCentral)

Two more definitions are useful:

Definition: A set is *closed* if it contains all its boundary points.

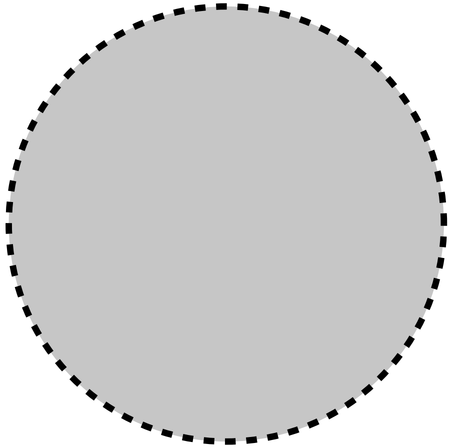
So sets that are defined by $\leq$, $\geq$ or $=$ signs are often closed.

Examples:



$\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$

is closed



$\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$

is not closed

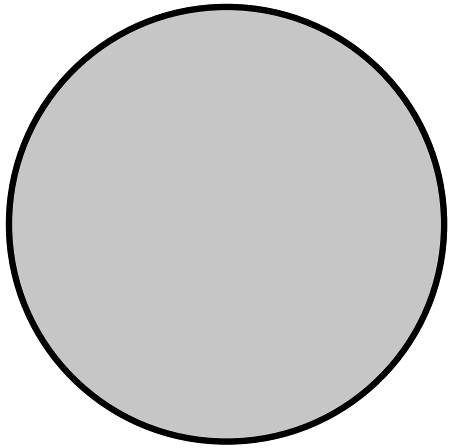
Two more definitions are useful:

Definition: A set is *closed* if it contains all its boundary points.
So sets that are defined by $\leq$, $\geq$ or $=$ signs are often closed.

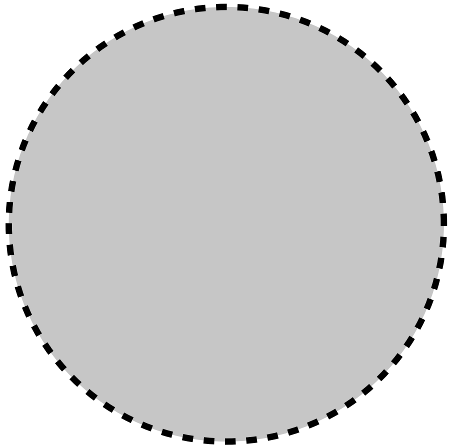
Definition: A set is *bounded* if it is contained in a large ball around the origin
Informally, a set is bounded if it doesn't "go to infinity".

Note that being bounded is unrelated to the boundary.

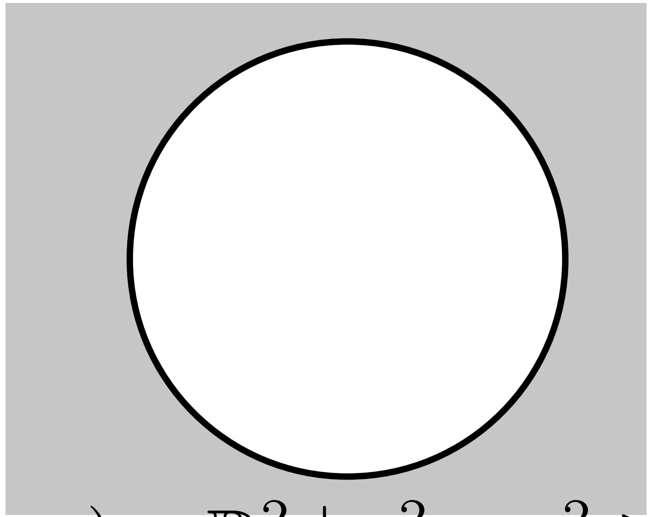
Examples:



$\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$
is closed and bounded



$\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$
is bounded but not closed



$\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \geq 1\}$
is closed but not bounded

A closed and bounded set in $\mathbb{R}^n$ behaves like a closed interval $[a, b]$ in $\mathbb{R}^1$, in the following sense:

Theorem 2: Continuous functions on closed and bounded sets have extremum: If f is a continuous function whose domain $\mathcal{D}$ is a closed and bounded set in $\mathbb{R}^n$, then there are points in $\mathcal{D}$ where f achieves its maximum and minimum values.

Note that f may have maximum and minimum values even if the domain is not closed or not bounded, see ex sheet #20 Q1d. P20-22 has some tips for determining whether extreme values exist when the domain is not closed or not bounded.

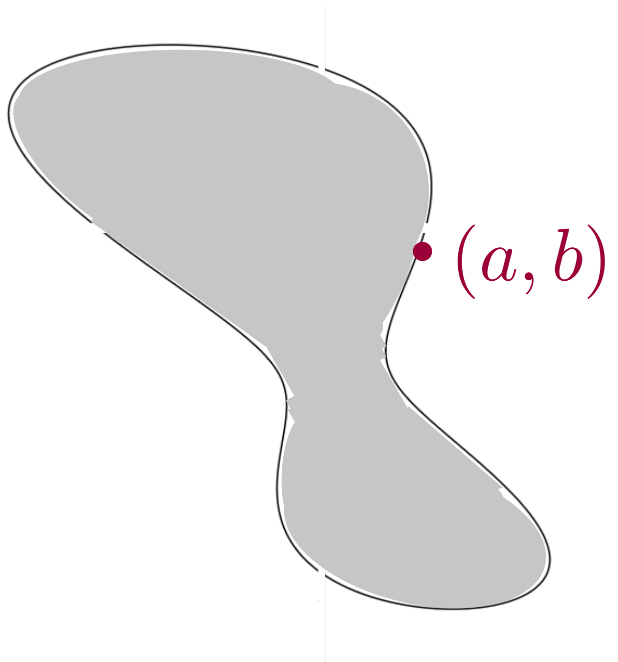
Remember our toy example of finding the tallest student by measuring all students and comparing these heights. This is the idea in the following algorithm.

Strategy for finding extreme values:

1. Draw the domain, and separate it into the interior (usually $<$ or $>$) and the boundary pieces (usually $=$).
2. Explain why the extremum exists, e.g. "f is a continuous function on a closed and bounded domain, so it has a maximum and a minimum."
3. Make a list of candidate extrema in the interior:
   - critical points: where $\nabla f = \mathbf{0}$.
   - singular points: where f is not differentiable.
4. Make a list of candidate extrema on each boundary piece: 2 possible methods
   - parametrise the boundary: for lines, boundaries of the form $y = p(x)$ or $z = p(x, y)$ where you can isolate one variable; circles and ellipses if f is simple.
   - Lagrange multipliers - circles and ellipses if f is complicated, other complicated boundaries
5. Compare the values of f at the candidate extrema from Steps 3 and 4.

Method 1 to find extrema on the boundary: boundary parametrisation

The main idea of this method is: suppose f achieves its maximum at (a, b), a point on the boundary. Then, if I were to walk along the boundary, I would experience a maximum of f when I pass through (a, b). In other words, if $t \to (x(t), y(t))$ is a parametrisation of a boundary piece, with $x(c) = a, y(c) = b$, then the single-variable function $f(x(t), y(t))$ must have a maximum at $t = c$ - i.e. c must be a critical point or singular point of $f(x(t), y(t))$.



This technique also works for boundary surfaces in $\mathbb{R}^3$: these are parametrised by $(s, t) \to (x(s, t), y(s, t), z(s, t))$, and we then look for critical and singular points of the 2-variable function $f(x(s, t), y(s, t), z(s, t))$ (see Homework 6 questions).

Example: Find the maximum value of $f(x, y) = x^2 + xy - 2y$ on the closed triangle with vertices $(0, 0)$, $(1, 0)$ and $(1, 1)$, and the point(s) where this maximum value is achieved.

Redo example: (ex. sheet #20 Q2) Find the maximum and minimum values of $f(x, y) = x^2 + y$ on the unit circle $x^2 + y^2 = 1$, and the point(s) where these extreme values are achieved.

The disadvantage of boundary parametrisation is that some implicit curves can be very hard to parametrise, e.g. $x^2y^2 + 1 = x^3 + y^5$. So we look at a second method to handle these cases, where we consider each boundary piece as (part of) a level set of a *constraint function* g.

Method 2 to find extrema on the boundary: Lagrange multipliers

Theorem 4: Lagrange Multipliers: Let f, g be n-variable functions, and suppose S is the set defined by $g(\mathbf{x}) = 0$ (i.e. the level set of g at $C = 0$). Suppose a point $\mathbf{a}$ in S is not an "endpoint" of S, and is a local maximum or minimum of f on S. If f and g have continuous first-order partial derivatives near $\mathbf{a}$ and $\nabla g(\mathbf{a}) \neq \mathbf{0}$, then $\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$ for some λ.

(The textbook states the conclusion as: there is a number λ_0 such that $(a_1, \ldots, a_n, \lambda_0)$ is a critical point of the *Lagrangian function*, defined to be $L(x_1, \ldots, x_n, \lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x})$. It is easy to show this is equivalent to our theorem statement above.)

Method 2 to find extrema on the boundary: Lagrange multipliers

Theorem 4: Lagrange Multipliers: Let f, g be n-variable functions, and suppose S is the set defined by $g(\mathbf{x}) = 0$ (i.e. the level set of g at $C = 0$). Suppose a point $\mathbf{a}$ in S is not an "endpoint" of S, and is a local maximum or minimum of f on S. If f and g have continuous first-order partial derivatives near $\mathbf{a}$ and $\nabla g(\mathbf{a}) \neq \mathbf{0}$, then $\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$ for some λ.

(The textbook states the conclusion as: there is a number λ_0 such that $(a_1, \ldots, a_n, \lambda_0)$ is a critical point of the *Lagrangian function*, defined to be $L(x_1, \ldots, x_n, \lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x})$. It is easy to show this is equivalent to our theorem statement above.)

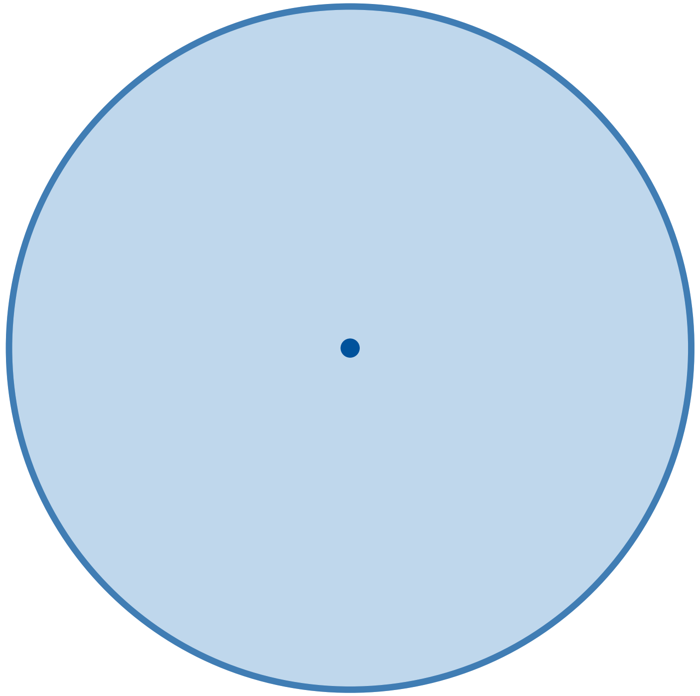
The theorem is about local extrema, but since every absolute extrema is a local extrema, it's also useful for finding absolute extrema. A rephrasing that's easier to use: If $\mathbf{a}$ is maximum or minimum of f on S, then one of the following is true:

- $\mathbf{a}$ is an "endpoint" of S (we can avoid this if S is defined by $<$, $>$ or $=$, not $\leq$ or $\geq$;
- f or g does not have continuous partial derivatives near $\mathbf{a}$;
- $\nabla g(\mathbf{a}) = \mathbf{0}$;
- $\nabla f(\mathbf{a}) = \lambda \nabla g(\mathbf{a})$ for some λ.

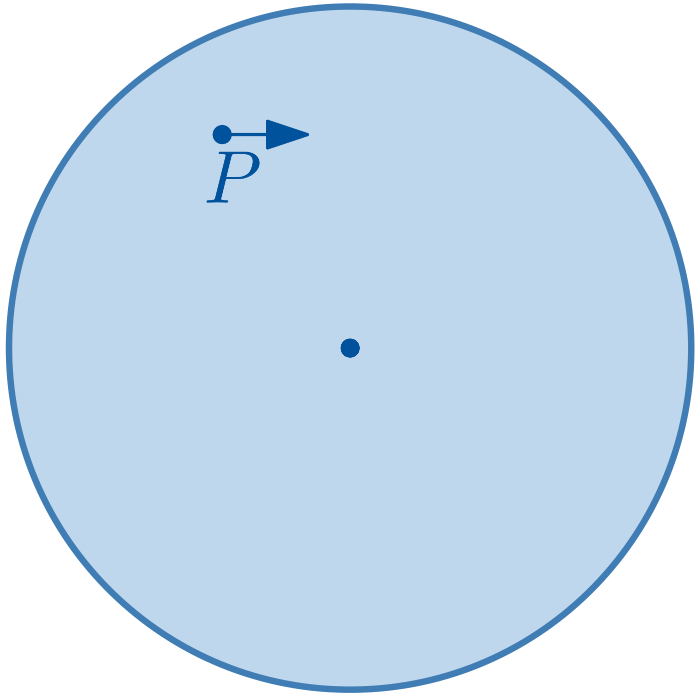
We will explain why Lagrange multipliers work after an example.

Redo example: (ex. sheet #20 Q2, p13) Use Lagrange multipliers to find the maximum and minimum values of $f(x, y) = x^2 + y$ on the unit circle $x^2 + y^2 = 1$, and the point(s) where these extreme values are achieved.

To see why the Lagrange multiplier technique works, consider the simple example of maximising $f(x, y) = x$ on the unit disk $\{x^2 + y^2 \leq 1\}$. And recall from p3 the view of ∇f as an "instruction" for how to maximise f (how to "climb higher").

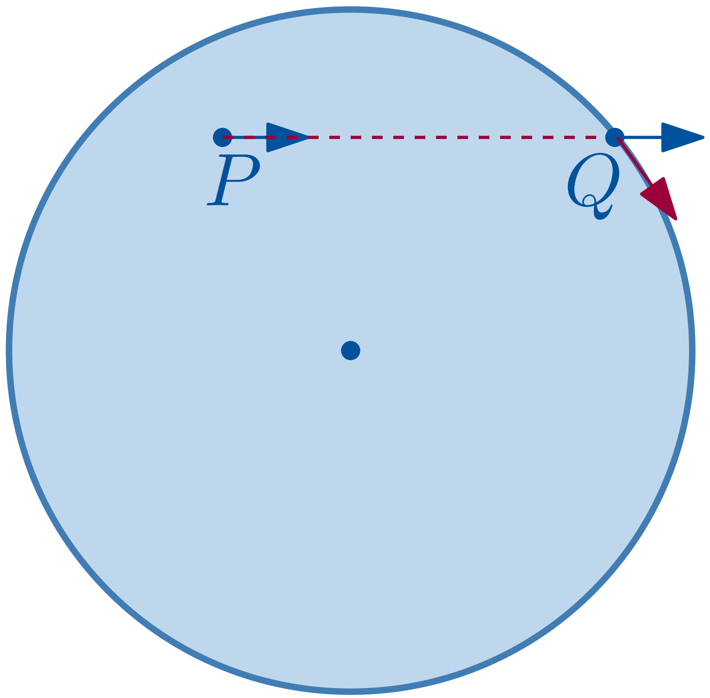


To see why the Lagrange multiplier technique works, consider the simple example of maximising $f(x, y) = x$ on the unit disk $\{x^2 + y^2 \leq 1\}$. And recall from p3 the view of ∇f as an "instruction" for how to maximise f (how to "climb higher").



Suppose I start at the point P. We have $\nabla f(P) = \mathbf{i}$ so I can walk in the $\mathbf{i}$ direction to increase f.

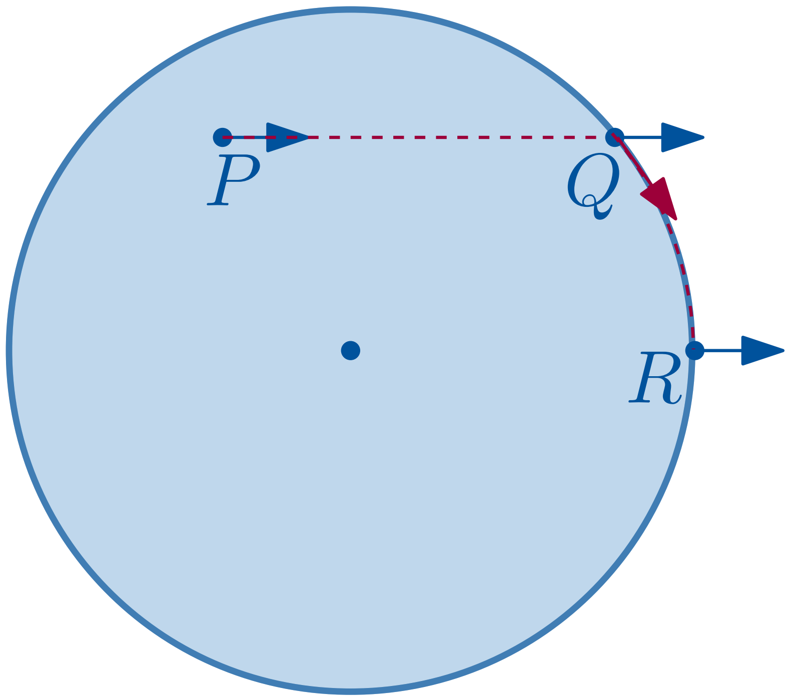
To see why the Lagrange multiplier technique works, consider the simple example of maximising $f(x, y) = x$ on the unit disk $\{x^2 + y^2 \leq 1\}$. And recall from p3 the view of ∇f as an "instruction" for how to maximise f (how to "climb higher").



Suppose I start at the point P. We have $\nabla f(P) = \mathbf{i}$ so I can walk in the $\mathbf{i}$ direction to increase f.

We have $\nabla f = \mathbf{i}$ at every point inside the disk, so following this instruction takes me to Q. At Q, I cannot continue in the direction $\nabla f(Q) = \mathbf{i}$, because it will take me outside the unit disk. However, there is a direction along the boundary (the red arrow) which makes an acute angle with $\nabla f(Q)$, so the directional derivative in this direction is positive. So I can still increase f by walking in this direction.

To see why the Lagrange multiplier technique works, consider the simple example of maximising $f(x,y) = x$ on the unit disk $\{x^2 + y^2 \leq 1\}$. And recall from p3 the view of ∇f as an "instruction" for how to maximise f (how to "climb higher").



Suppose I start at the point P. We have $\nabla f(P) = \mathbf{i}$ so I can walk in the $\mathbf{i}$ direction to increase f.

We have $\nabla f = \mathbf{i}$ at every point inside the disk, so following this instruction takes me to Q. At Q, I cannot continue in the direction $\nabla f(Q) = \mathbf{i}$, because it will take me outside the unit disk. However, there is a direction along the boundary (the red arrow) which makes an acute angle with $\nabla f(Q)$, so the directional derivative in this direction is positive. So I can still increase f by walking in this direction.

This works until I reach R. At R, there are no directions along the boundary which make an acute angle with $\nabla f(R) = \mathbf{i}$, because the boundary is perpendicular to $\mathbf{i}$.

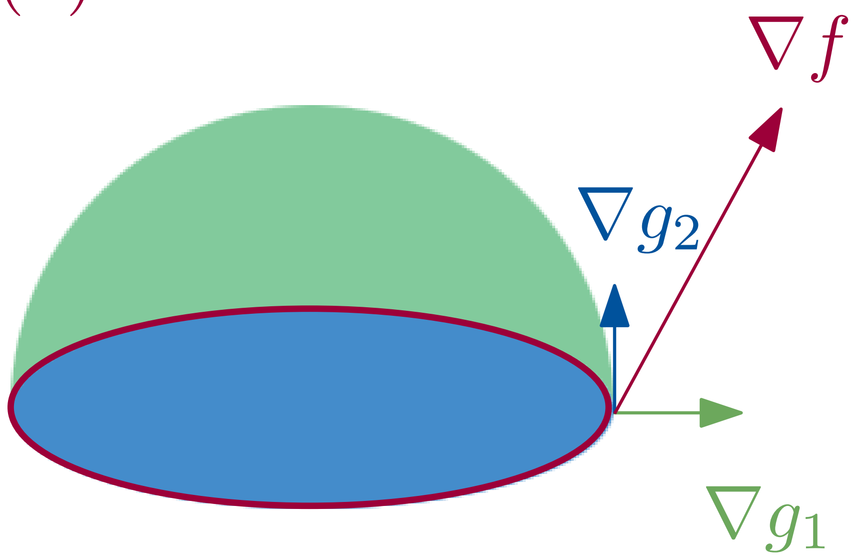
So, at an extremum on the boundary, the boundary is perpendicular to ∇f. If the boundary is the level set of g, then the boundary is perpendicular to ∇g (assuming $\nabla g \neq \mathbf{0}$). So ∇f and ∇g must be in the same direction, i.e. $\nabla f = \lambda \nabla g$.

Non-examinable: Lagrange multipliers with multiple constraints

Sometimes regions have boundary pieces S that are defined by more than one equation - i.e. S is the intersection of the level sets of multiple constraint functions $g_1, g_2, \dots, g_k$. (e.g. the red "edge" of the upper half-ball, p7: $g_1(x, y, z) = x^2 + y^2 + z^2 - 1$, $g_2(x, y, z) = z$.)

In this case, the boundary piece S is perpendicular to ∇g_i for all i (assuming that $\{\nabla g_i\}$ is linearly independent, which will almost always be the case). And if $\mathbf{a}$ achieves an extremum of f on S, then $\nabla f(\mathbf{a})$ must be perpendicular to S. So the necessary condition is $\nabla f(\mathbf{a}) = \lambda_1 \nabla g_1(\mathbf{a}) + \dots + \lambda_k \nabla g_k(\mathbf{a})$.

This technique is non-examinable. If you need to find extrema on a set defined by more than one equation, you will be able to use other methods e.g. parametrisation (see Homework 6 final question). You can also choose to use this method if you prefer.



(picture from URegina MathCentral)

Lagrange multipliers or boundary parametrisation?

We have seen two different ways to find candidate extrema on the boundary. In many examples, both methods work, and even for the same situation different people may have different opinions about which method they prefer. You should try both methods on some of the homework questions or practice questions, to find out which method is easier for you in different examples. My preferences are:

- If the boundary has an easy parametrisation (e.g. lines, planes, other cases where you can easily solve for one variable), I would choose parametrisation.
- If the boundary is a circle or ellipse in $\mathbb{R}^2$, I look at the objective function:
  - If the objective function is a low degree polynomial, I try parametrisation first.
  - Otherwise, I try Lagrange multipliers first.
- For other sets I will generally use Lagrange multipliers.

If my chosen method leads to equations that are very difficult to solve, I will try the other method to see if it leads to easier equations.

Note that boundary parametrisation reduces the number of variables, whereas Lagrange multipliers increases the number of variables (so the conditions are generally harder to solve).

Extrema on domains that are not bounded or not closed

If we wish to maximise a function f on a domain that is not bounded or not closed, then there are two possibilities:

- There is a maximum (p22). We can either find it directly using inequalities, or we apply the general strategy (p10) to find a possible maximum $f(\mathbf{a})$, and then using inequalities to show that $f(\mathbf{x}) \leq f(\mathbf{a})$ for all $\mathbf{x}$ in the domain. (You can also use these non-algorithmic inequality methods on closed and bounded domains.)
- There is no maximum (p21). We can either prove this by considering the behaviour of f "at infinity" or near a part of the boundary that is not in the domain, or by applying the general strategy (p10) and then showing that none of the candidate points can be a maximum.

The case where there is no extremum is usually easier to prove. If you already have information about the function from earlier parts of the question, this can be helpful.

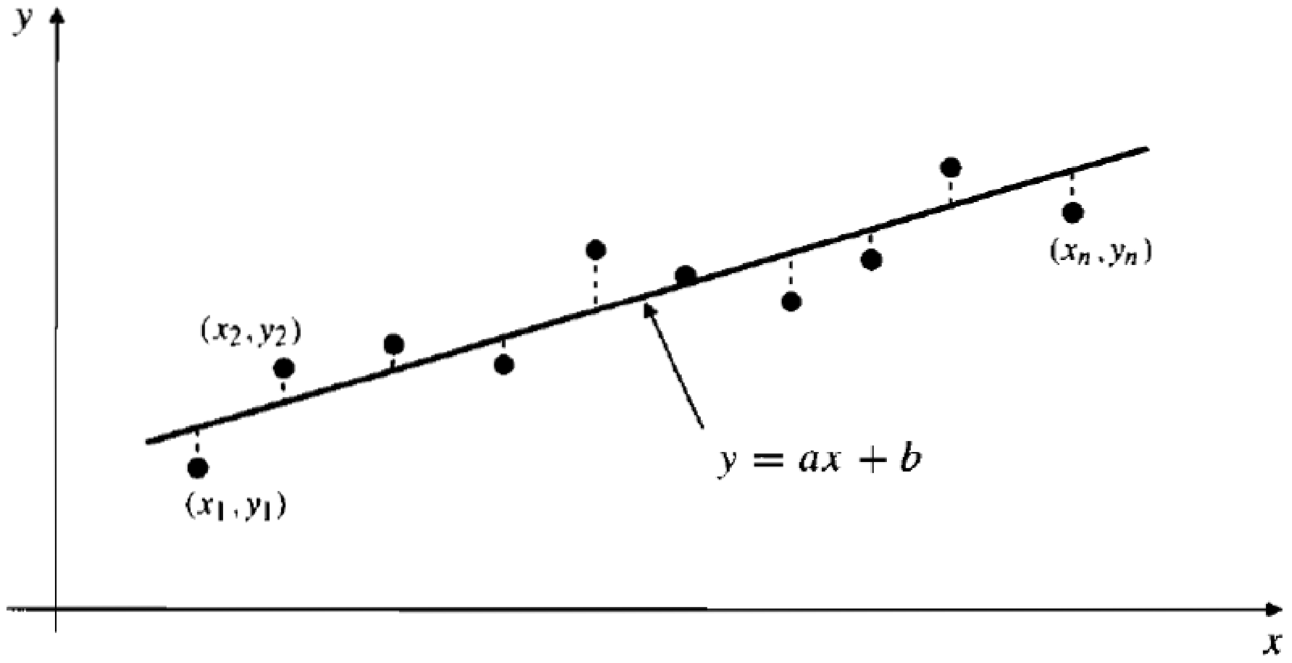
Example: Does $f(x, y) = 2x^3 + 6xy + 3y^2$ have a maximum on $\mathbb{R}^2$? (You found on ex. sheet #19 Q1 that $(1, -1)$ is a local minimum and $(0, 0)$ is a saddle point, and there are no other critical points.)

Be careful: a function that does not "go to infinity" might still not have a maximum or minimum.

Example: Does $f(x, y) = \dfrac{1}{1 + x^2 + y^2}$ have a maximum and a minimum on $\mathbb{R}^2$?

Non-examinable: linear regression as an application of extremisation

Suppose two physical quantities x and y (e.g. temperature and pressure) are related by $y = ax + b$, for some unknown constants a and b. To estimate a and b, we can do an experiment to find some data $(x_1, y_1), \ldots, (x_n, y_n)$, then plot the data and draw a line that "best" fits the data. Mathematically, this means we want to maximise (or minimise) some function $f(a, b)$, where f measures "how well (or how badly) the line fits the data" - what this means will depend on the physical situation.



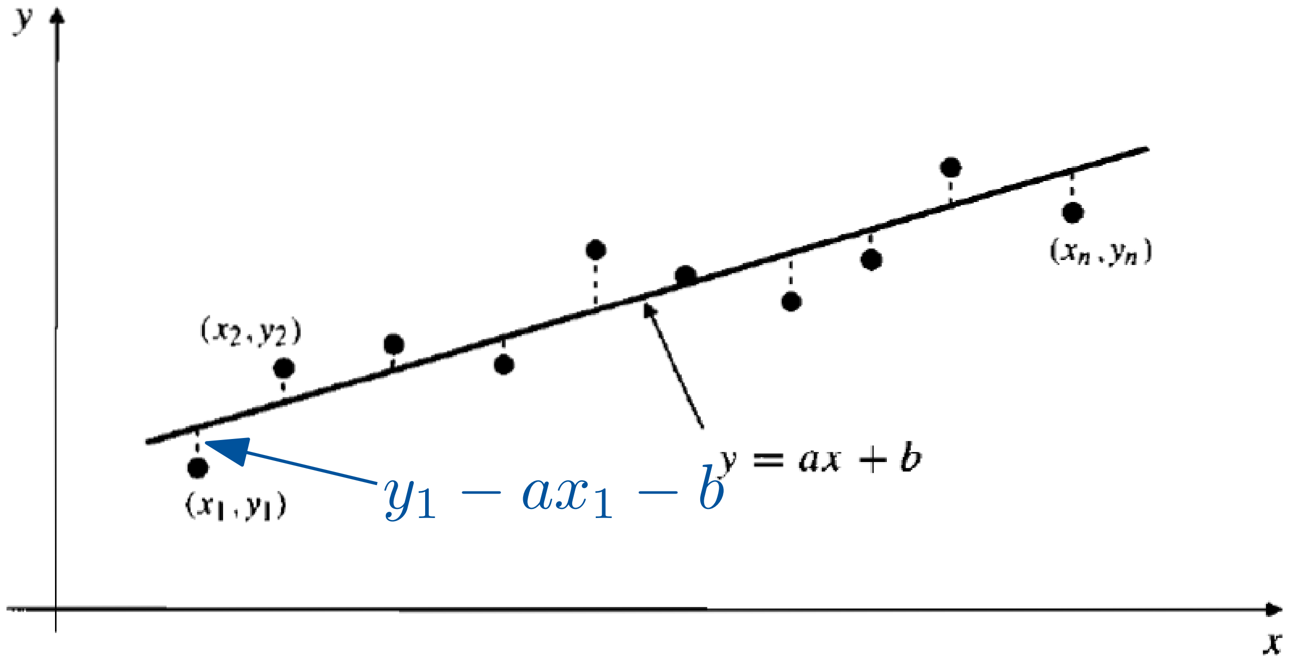
Non-examinable: linear regression as an application of extremisation

Suppose two physical quantities x and y (e.g. temperature and pressure) are related by $y = ax + b$, for some unknown constants a and b. To estimate a and b, we can do an experiment to find some data $(x_1, y_1), \ldots, (x_n, y_n)$, then plot the data and draw a line that "best" fits the data. Mathematically, this means we want to maximise (or minimise) some function $f(a, b)$, where f measures "how well (or how badly) the line fits the data" - what this means will depend on the physical situation.

One common and convenient scheme is least squares, where we minimise the error function

$$f(a,b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$

the sum of squares of the vertical distances from the data points to the line (dotted lines in the diagram).



Reminder: given data points $(x_1, y_1), \ldots, (x_n, y_n)$, we wish to minimise the error function $f(a,b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2$.

(The notation is confusing: a, b are the unknowns, and x_i, y_i are known numbers.)

The domain of f (i.e. the possible values of (a,b)) is all of $\mathbb{R}^2$, which has no boundary. So, if a minimum for f exists, it must be at a critical point (because f is differentiable everywhere, so it has no singular points).

At a critical point:

$$\frac{\partial f}{\partial a} = 0 \Rightarrow \sum_{i=1}^{n} 2(y_i - ax_i - b)(-x_i) = 0 \Rightarrow \left(\sum_{i=1}^{n} x_i^2\right) a + \left(\sum_{i=1}^{n} x_i\right) b = \sum_{i=1}^{n} x_i y_i$$

$$\frac{\partial f}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} 2(y_i - ax_i - b)(-1) = 0 \Rightarrow \left(\sum_{i=1}^{n} x_i\right) a + \left(\sum_{i=1}^{n} 1\right) b = \sum_{i=1}^{n} y_i$$

Divide each equation on the far right hand side by n, and use the mean value notation $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$:

$$\overline{x^2}a + \overline{x}b = \overline{xy}$$

$$\overline{x}a + b = \overline{y}$$

Combine into a matrix equation:

$$\begin{bmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \overline{xy} \\ \overline{y} \end{bmatrix}$$

So $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \overline{xy} \\ \overline{y} \end{bmatrix}$, which means $a = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2}$; $\quad b = \dfrac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - (\overline{x})^2}$.

Divide each equation on the far right hand side by n, and use the mean value notation $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$:

$$\overline{x^2}a + \overline{x}b = \overline{xy}$$
$$\overline{x}a + b = \overline{y}$$

Combine into a matrix equation:

$$\begin{bmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \overline{xy} \\ \overline{y} \end{bmatrix}$$

So $\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \overline{xy} \\ \overline{y} \end{bmatrix}$, which means $a = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2}$; $\quad b = \dfrac{\overline{x^2}\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - (\overline{x})^2}$.

To conclude that these values of (a, b) really minimises $f(a, b) = \displaystyle\sum_{i=1}^{n}(y_i - ax_i - b)^2$, we need to show that f achieves a minimum on $\mathbb{R}^2$. A precise proof is complicated. The main idea: as (a, b) "moves away from the origin", $f(a, b) \to \infty$ (and $\mathbb{R}^2$ has no boundary so we do not need to consider (a, b) moving towards a boundary point that is not in the domain).