28 Max-min problems, least squares

In this topic we will look at simple max-min problems. As with single variable functions, we will spot critical points by looking where the derivatives are 0. Continuing the analogy, we will distinguish maxima from minima using a second derivative test. Since, with multiple variables, there are multiple first and second derivatives, our work will be slightly more complicated than in 18.01.

28.1 Standard calculus question

Given a function \( z = f(x, y) \), where are the (relative) maxima and minima?

**Answer:** At a relative maximum or minimum we have \( \nabla f = \langle \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \rangle = \langle 0, 0 \rangle \), i.e. the first derivatives \( \frac{\partial f}{\partial x} = 0 \) and \( \frac{\partial f}{\partial y} = 0 \).

**Geometric justification.** The basic geometric reason is that at a maximum (or minimum) the tangent plane must be horizontal. That is, its normal must be in the \( k \) direction, i.e. parallel to \( \langle 0, 0, 1 \rangle \). But, we know the tangent plane at \( (x_0, y_0, z_0) \) has equation

\[
z - z_0 = \left. \frac{\partial f}{\partial x} \right|_o (x - x_0) + \left. \frac{\partial f}{\partial y} \right|_o (y - y_0),
\]

i.e. it has normal \( \langle -\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1 \rangle \). Comparing this with \( \langle 0, 0, 1 \rangle \) we see that, for the plane to be horizontal, the partials must be 0. –See the examples below.

**Critical points:** Points where both partial derivatives are 0 are called critical points.

**Example 28.1.** Consider \( z = 4 - x^2 - y^2 \). Find the critical points. Use a sketch to decide if they are maxima, minima or neither.

**answer:** The gradient \( \nabla z = \langle -2x, -2y \rangle \). So, the only critical point is \( (0, 0) \). From the sketch, we see this is clearly a maximum.

Sketch of \( z = 4 - x^2 - y^2 \) with horizontal tangent plane at \( (0,0,4) \).
This surface is called an elliptic paraboloid.

**Example 28.2.** Consider \( z = 1 + x^2 + y^2 \). Find the critical points. Use a sketch to decide if they are maxima, minima or neither.

**answer:** The gradient \( \nabla z = \langle 2x, 2y \rangle \). So, the only critical point is \((0, 0)\). From the sketch, we see this is clearly a minimum.

Sketch of \( z = 1 + x^2 + y^2 \) with horizontal tangent plane at \((0, 0, 1)\).

This surface is called an elliptic paraboloid.

**Example 28.3.** Consider \( z = y^2 - x^2 \). Find the critical points. Use a sketch to decide if they are maxima, minima or neither.

**answer:** The gradient \( \nabla z = \langle -2x, 2y \rangle \). So, the only critical point is \((0, 0)\). From the sketch, we see this is clearly neither a maximum or minimum. It is called a saddle point.

Sketch of \( z = y^2 - x^2 \) with horizontal tangent plane at \((0, 0, 0)\).

This surface is called a hyperbolic paraboloid. Notice that the level curves at \( z = 0 \) are two intersecting lines.

**Example 28.4.** Making a box. Suppose we need to make a cardboard box with the following specifications:

- The front and back are made from a single sheet of cardboard.
- The sides are double thick.
- The bottom is triple thick. There is no top.
- The volume = 3.

What dimensions use the least amount of cardboard?
answer: We start by finding the total area of the cardboard needed.

The area of one side = \( yz \). There two double thick sides, so the cardboard used for the sides is \( 4yz \).

The area of the front (and back) = \( xz \). These are single thick, so the cardboard used for the front and back is \( 2xz \).

The area of the bottom = \( xy \). It is triple thick, so the cardboard used for the bottom is \( 3xy \).

Adding these together we get: total cardboard used = \( A = 4yz + 2xz + 3xy \).

We are give that the volume = 3 = \( xyz \). We use this to eliminate \( z \) from the formula for \( A \): \( z = \frac{3}{xy} \), so

\[
A = 4y \cdot \frac{3}{xy} + 2x \cdot \frac{3}{xy} + 3xy = \frac{12}{x} + \frac{6}{y} + 3xy.
\]

To find the critical points we compute the partial derivatives and set them to 0

\[
A_x = -\frac{12}{x^2} + 3y = 0
\]
\[
A_y = -\frac{6}{y^2} + 3x = 0.
\]

These are two equations in two unknowns. We solve by eliminating \( y \).

The first equation implies \( y = \frac{4}{x^2} \). Putting this in the second equation gives

\[
-\frac{6}{16} x^4 + 3x = 0
\]

This factors as \( 3x(1 - x^3/8) = 0 \), so \( x = 0 \) or \( x = 2 \). We reject \( x = 0 \) because it is not physically meaningful (also the formula for area would give infinity), so \( x = 2 \). Using our formulas for \( y \) and \( z \) we have \( x = 2, \ y = 1, \ z = 3/2 \).

Later we will see how to check that this gives a minimum using a second derivative test. Here we’ll just accept that physically there must be a minimum for some positive \( (x, y, z) \). Since there is only one such critical point, it must be the minimum.

28.2 Least squares

As an application of finding maxima and minima we will look at the least squares fitting of a curve to data.
A standard practice in science is to fit a curve to data points. Often we require the curve to be a line, but other curves are also used.

**Example 28.5.** Plot the data points (0,1), (1,1), (2,3), (3,2) in the $xy$-plane and guess at the line that best fits the data.

**answer:** Guessing is easy, how about

Of course, it’s impossible to draw a line that goes through all the points exactly, so we tried to draw one that ‘fits’ the data as best we could.

### 28.2.1 The least squares model

Suppose we want to see how age affects vision. A standard experiment would be to select a random group of subjects and ask them their age and measure their visual acuity. We would collect data in the form of pairs of numbers

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n),$$

where $x$ is the subject’s age and $y$ their visual acuity.

A reasonable goal of our experiment would be to find a formula relating $x$ and $y$, i.e. $y = f(x)$. That way, if I told you someone’s age, you could use the formula to predict their visual acuity. The error in the model would be the difference between your prediction and their actual acuity.

In this model, $x$ is called the independent or predictive or explanatory variable and visual acuity is called the dependent or response variable.

Least squares is a method to minimize the overall error between the actual data and the fitted line. It is called least squares because it minimizes the sum of the individual errors squared. There are good statistical reasons for doing this.

Least square is sometimes called linear regression. Be careful, the term linear does not refer to fitting a line to data. Linear regression can be used to fit a parabola, a cubic, an exponential, etc.

### 28.3 The least squares method for fitting a line to data

Start with $n$ data points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

We want to find the line $y = ax + b$ that best fits the data. That is, we need to find the best values of $a$ and $b$. 
Imagine we have a candidate line \( y = ax + b \). For each data point \((x_i, y_i)\), the model predicts \( y = ax_i + b \), and the data says that \( y = y_i \). The error between the actual and predicted value is

\[
\text{error}_i = y_i - (ax_i + b)
\]

In least squares we consider the sum of these errors squared:

\[
E(a, b) = \sum_{i=1}^{n} \text{error}_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2.
\]

This is a measure of the total error and depends on \( a \) and \( b \). Our goal is to find the values of \( a \), \( b \) that minimize \( E(a, b) \).

Finally, we have a calculus problem! We take the partial derivatives and set them to 0.

\[
\frac{\partial E}{\partial a} = \sum_{i=1}^{n} -2x_i(y_i - ax_i - b) = 2 \sum_{i=1}^{n} (ax_i^2 + bx_i - x_i y_i) = 0.
\]

\[
\frac{\partial E}{\partial b} = \sum_{i=1}^{n} -2(y_i - ax_i - b) = 2 \sum_{i=1}^{n} (ax_i + b - y_i) = 0.
\]

We can collect these into the least squares equations for the best fitting line. (Remember, the unknowns are \( a \) and \( b \).)

\[
\begin{align*}
\left( \sum_{i=1}^{n} x_i^2 \right) a + \left( \sum_{i=1}^{n} x_i \right) b &= \sum_{i=1}^{n} x_i y_i \\
\left( \sum_{i=1}^{n} x_i \right) a + nb &= \sum_{i=1}^{n} y_i
\end{align*}
\]

**Example 28.6.** Use least squares to find the best fitting line to the data points (0,1), (1,1), (2,3), (3,2).

**answer:** We have to compute all the coefficients in the least squares equations. There are 4 data points, so \( n = 4 \).

\[
\begin{align*}
\sum_{i=1}^{4} x_i^2 &= 0^2 + 1^2 + 2^2 + 3^2 = 14, & \sum_{i=1}^{4} x_i &= 0 + 1 + 2 + 3 = 6 \\
\sum_{i=1}^{4} x_i y_i &= 0 + 1 + 6 + 6 = 13, & \sum_{i=1}^{4} y_i &= 1 + 1 + 3 + 2 = 7
\end{align*}
\]

The least squares equations are

\[
14a + 6b = 13 \\
6a + 4b = 7
\]

We solve this in matrix form

\[
\begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 13 \\ 7 \end{bmatrix} \Rightarrow \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 13 \\ 7 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 4 & -6 \\ -6 & 14 \end{bmatrix} \begin{bmatrix} 13 \\ 7 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.0 \end{bmatrix}
\]

Our best fitting line is \( y = 0.5x + 1.0 \).

28.3.1 Least squares fitting a parabola

The same logic allows us to fit other curves to data. Note this is still called linear regression!

**Example 28.7.** For the same points as in the previous example use least squares to find the best fitting parabola.

**answer:** The logic is the same as for a line. We just use a quadratic to to predict the value of \( y \).

A parabola has equation \( y = ax^2 + bx + c \). For a data point \((x_i, y_i)\) the error between the actual and predicted value of \( y \) is

\[
\text{error}_i = y_i - (ax_i^2 + bx_i + c).
\]

The total squared error is a function of \( a, b \) and \( c \):

\[
E(a, b, c) = \sum \text{error}_i^2 = \sum (y_i - ax_i^2 - bx_i - c)^2.
\]

For the least squares fit we need to minimize \( E(a, b, c) \). As usual, we set partial derivatives to 0.

\[
\begin{align*}
\frac{\partial E}{\partial a} &= \sum -2x_i^2(y_i - ax_i^2 - bx_i - c) = 0 \\
\frac{\partial E}{\partial b} &= \sum -2x_i(y_i - ax_i^2 - bx_i - c) = 0 \\
\frac{\partial E}{\partial c} &= \sum -2(y_i - ax_i^2 - bx_i - c) = 0
\end{align*}
\]

Collecting up terms we have the least squares equations for the best fitting parabola

\[
\begin{align*}
\left( \sum_{i=1}^{n} x_i^4 \right) a + \left( \sum_{i=1}^{n} x_i^3 \right) b + \left( \sum_{i=1}^{n} x_i^2 \right) c &= \sum_{i=1}^{n} x_i^2 y_i \\
\left( \sum_{i=1}^{n} x_i^3 \right) a + \left( \sum_{i=1}^{n} x_i^2 \right) b + \left( \sum_{i=1}^{n} x_i \right) c &= \sum_{i=1}^{n} x_i y_i \\
\left( \sum_{i=1}^{n} x_i^2 \right) a + \left( \sum_{i=1}^{n} x_i \right) b + nc &= \sum_{i=1}^{n} y_i
\end{align*}
\]
Plugging in the data in our example:

\[
\begin{align*}
\sum x_i^4 &= 98, & \sum x_i^3 &= 36, & \sum x_i^2 &= 14, & \sum x_i &= 6 \\
\sum x_i^2y_i &= 31, & \sum x_iy_i &= 13, & \sum y_i &= 7
\end{align*}
\]

So our least squares equations are

\[
\begin{align*}
98a + 36b + 14c &= 31 \\
36a + 14b + 6c &= 13 \\
14a + 6b + 4c &= 7
\end{align*}
\]

Solving (we used Octave) we get \(a = -0.25, \ b = 1.25, \ c = 0.75\)

Best fitting parabola to data