About analytical optimisation

Frank Miller, Department of Computer and Information Science, Linköpings University

Spring 2023

Analytical one-dimensional optimisation

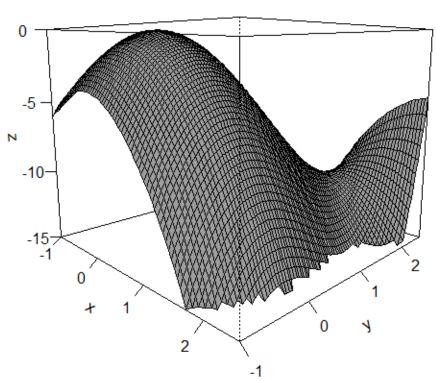
If we want to maximise a one-dimensional function, for example $g(x) = 4 + x - x^2$, we use the first and second derivative. We set the first derivative to 0 and solve the equation. Solutions are then investigated with the second derivative: if it is negative, we have found a maximum; if it is positive, we have found a minimum; if it is 0, we cannot be sure what it is and have to do further investigations.

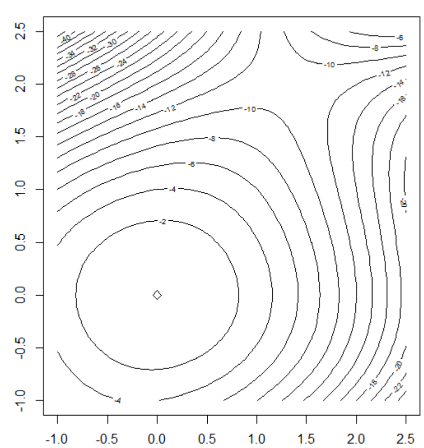
Example for analytical two-dimensional optimisation

Suppose we want to determine the values x and y such that the following function becomes maximal:

$$g(x, y) = -3x^2 - 4y^2 + xy^3.$$

In this case, it is possible to calculate these values analytically. In the left figure, you can see a 3d-plot of this function (where $z = g(x, y)$). In the right figure, you can see a contour plot of it with x and y at the two axis and the function value shown in terms of contours.





Here, we have a two-dimensional case, but we can generalise the computation from the one-dimensional case. Corresponding to the first derivative is the gradient, corresponding to the second derivative is the Hessian matrix. We compute them now for this example.

Gradient

The gradient is a vector; each component is the derivative with respect to one variable. The derivative with respect to x is $-6x + y^3$ and with respect to y it is $-8y + 3xy^2$. The gradient is therefore:

$$g'(\begin{pmatrix} x \\ y \end{pmatrix}) = \begin{pmatrix} -6x + y^3 \\ -8y + 3xy^2 \end{pmatrix}.$$

The gradient at a point $(x_0, y_0)^\top$ can be interpreted as the direction of steepest increase of g in this point.

Hessian

The Hessian matrix is the collection of second order derivatives. Here, we have the second derivative with respect to x (-6), the second derivative with respect to y $(3y^2)$, and the derivative with respect to x and then to y $(-8 + 6xy)$. The Hessian matrix is then

$$g''(\begin{pmatrix} x \\ y \end{pmatrix}) = \begin{pmatrix} -6 & 3y^2 \\ 3y^2 & -8 + 6xy \end{pmatrix}.$$

The Hessian matrix at a point $(x_0, y_0)^\top$ gives information about the local curvature of g in this point.

Set gradient to 0

We get two equations, $-6x + y^3 = 0$ and $-8y + 3xy^2 = 0$. The first gives $x = y^3/6$ which we plug in into the second: $8y = y^5/2$. So $y = 0$ or $16 = y^4$. This gives three possibilities for y: $y = -2, 0, 2$. Using $x = y^3/6$, we identify the following three points where the gradient is the 0-vector:

$$\begin{pmatrix} -4/3 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4/3 \\ 2 \end{pmatrix}.$$

Investigate the Hessian matrix

We compute the Hessian matrix for the second and the third point (the first point is similar to the third):

$$g''(\begin{pmatrix} 0 \\ 0 \end{pmatrix}) = \begin{pmatrix} -6 & 0 \\ 0 & -8 \end{pmatrix}.$$

One can check that the condition for negative definitness is fulfilled for this matrix and consequently, we have shown that we have a local maximum at $(0, 0)^\top$.

$$g''(\begin{pmatrix} 4/3 \\ 2 \end{pmatrix}) = \begin{pmatrix} -6 & 12 \\ 12 & 8 \end{pmatrix}.$$

The eigenvalues of this matrix are $-12.89, 14.89$ (they can be computed analytically as solutions for λ in $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and one obtains then $1 \pm \sqrt{193}$; you can check the result with the `R`-function `eigen`). Since one eigenvalue is negative, the other positive, the Hessian matrix is indefinite, and the point $(4/3, 2)^\top$ is a saddle point of g.

The results found here analytically can be confirmed in the figure above.

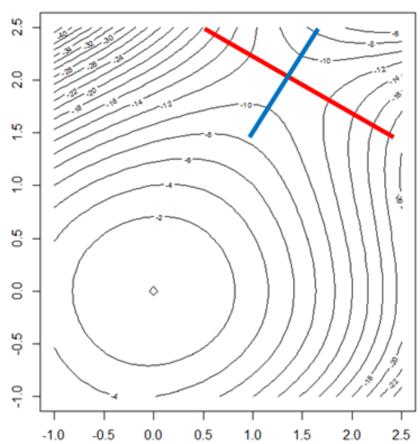
The eigenvectors

Each eigenvalue has an eigenvector associated. We get even more insight about a saddle point (or maximum, minimum) if we consider eigenvectors. If one moves in the direction of the eigenvector through a point $\mathbf{x}_0$ where the gradient of g is 0, the eigenvalue can be interpreted as second derivative in that direction.

In our example, we consider the saddle point $\mathbf{x}_0 = (4/3, 2)^\top$. The eigenvectors are

$$\begin{pmatrix} -0.867 \\ 0.498 \end{pmatrix}, \quad \begin{pmatrix} 0.498 \\ 0.867 \end{pmatrix},$$

for the eigenvalues $-12.89, 14.89$, respectively (in red and blue, respectively, in the figure; $\mathbf{x}_0 = (4/3, 2)^\top$ is where the two lines cross). Therefore, going through $\mathbf{x}_0$ into the direction of the first eigenvector (red) means that $\mathbf{x}_0$ is a local maximum in this direction; whereas it is a local minimum in the direction of the second eigenvector (blue).



Definitions and results about definite symmetric matrices

Let A be a symmetric $n \times n$-matrix ($A^\top = A$). Then:

- A is called **positive definite** if $\mathbf{x}^\top A\mathbf{x} > 0$ for all n-dimensional vectors $\mathbf{x} \neq \mathbf{0}$. This is fulfilled if and only if all n eigenvalues are positive.
- A is called **negative definite** if $\mathbf{x}^\top A\mathbf{x} < 0$ for all n-dimensional vectors $\mathbf{x} \neq \mathbf{0}$. This is fulfilled if and only if all n eigenvalues are negative.
- A is called **positive semi-definite** if $\mathbf{x}^\top A\mathbf{x} \geq 0$ for all n-dimensional vectors $\mathbf{x}$. This is fulfilled if and only if all n eigenvalues are ≥ 0.
- A is called **negative semi-definite** if $\mathbf{x}^\top A\mathbf{x} \leq 0$ for all n-dimensional vectors $\mathbf{x}$. This is fulfilled if and only if all n eigenvalues are ≤ 0.
- A is called **indefinite** if $\mathbf{x}_1^\top A\mathbf{x}_1 > 0$ and $\mathbf{x}_2^\top A\mathbf{x}_2 < 0$ for two n-dimensional vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. This is fulfilled if and only if at least one eigenvalue is positive and at least one eigenvalue is negative.

Results on conditions for maximum, minimum, and saddle point

Let $g : I\!R^n \rightarrow I\!R$ be an at least two times continuously differentiable function. Let $\mathbf{x}_0$ be a vector where the gradient of g is $\mathbf{0}$. Then:

- $\mathbf{x}_0$ is a local maximum if the Hessian matrix at $\mathbf{x}_0$ is negative definite.
- $\mathbf{x}_0$ is a local minimum if the Hessian matrix at $\mathbf{x}_0$ is positive definite.
- $\mathbf{x}_0$ is a saddle point if the Hessian matrix at $\mathbf{x}_0$ is indefinite.

Notation

Instead of writing $g'(\mathbf{x})$ for the gradient and $g''(\mathbf{x})$ for the Hessian, the notation $\nabla g(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$ is often used in literature.