CHEM1034 - Week 25 Lecture - Multivariate optimisation

Chapters 21, 22 of Monk and Munro, "Maths for Chemistry", 2nd edition.
Section 4.10 of Steiner, "The Chemistry Maths Book", 2nd edition.

A frequent requirement in physical sciences, engineering and economics is to find a set of parameters of a physical system, device or arrangement of assets that maximizes or minimizes a certain function. For example the yield of a chemical reaction chain might depend on concentrations, pressures, temperatures and other parameters and we are interested in maximising the yield and minimising the amount of unproductive reagent loss. Yield and loss are functions of all those concentrations, temperatures and pressures:

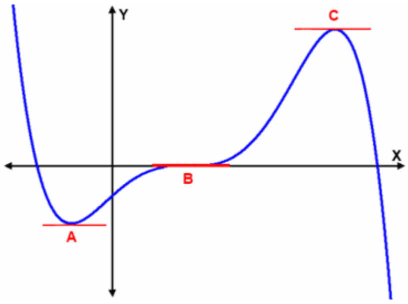


Figure 1. *A local minimum (A), an inflection point (B) and a local maximum (C) of a univariate function.*

$$\text{yield} = f(x, y, z, \ldots) \qquad \text{loss} = g(x, y, z, \ldots) \tag{1}$$

and we need find the *optimal* values of the parameters that would maximize the yield-to-loss ratio:

$$\{x_{\text{opt}}, y_{\text{opt}}, z_{\text{opt}}, \ldots\} = \underset{x,y,z,\ldots}{\arg\max} \frac{f(x, y, z, \ldots)}{g(x, y, z, \ldots)} \tag{2}$$

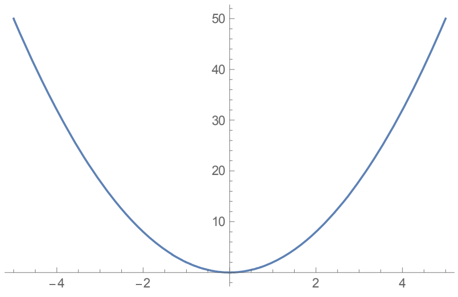
Such problems are known as *optimization problems*.

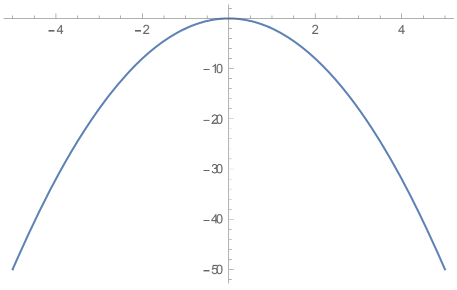
Reminder: single variable case

In the case of continuous functions of one variable they involve identification of points in which first derivatives are zero – in those points the tangent line is horizontal (Figure 1), which can only happen at a minimum (A), maximum (C) or an inflection point (B).

The nature of the point may be determined by computing the second derivative. If the second derivative is positive, we have a minimum, if it is negative we have a maximum. This is most clearly illustrated by a pair of quadratic functions:

$$f(x) = 2x^2 \qquad\qquad g(x) = -2x^2$$





First derivatives of both functions are zero at the origin:

$$\left.\frac{df(x)}{dx}\right|_{x=0} = 0 \qquad \left.\frac{dg(x)}{dx}\right|_{x=0} = 0$$

Second derivatives are positive in the case of a minimum and negative in the case of a maximum:

$$\left.\frac{d^2 f(x)}{dx^2}\right|_{x=0} = 4 \qquad \left.\frac{d^2 g(x)}{dx^2}\right|_{x=0} = -4$$

Zero second derivatives indicate *the possibility* of an inflection point, but may also indicate very shallow (cubic, quartic or higher order) minima or maxima. A few examples are given in Figure 2.

$f(x) = x^3$ $\qquad$ $g(x) = x^4$ $\qquad$ $h(x) = -x^4$

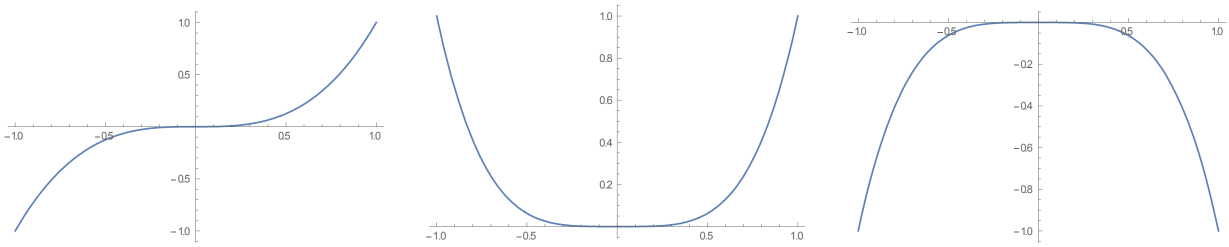


Figure 2. *Examples of stationary points in which both the first and the second derivatives of the function are zero, illustrating the fact that, when the second derivative is zero at a particular stationary point, further investigation is normally required.*

As a general recipe, to find the stationary points (also called *turning points*) of a function $f(x)$:

1. Differentiate the function to obtain $f'(x)$.
2. Find *all* solutions of $f'(x) = 0$.
3. Differentiate the function again to obtain $f''(x)$.
4. Substitute the solutions of $f'(x) = 0$ into $f''(x)$ and find out its sign.

<u>Example 1</u>: by considering its derivative, determine the intervals over which $f(x) = x^3 - 6x^2 + 9x + 1$ is increasing and the intervals over which it is decreasing. Determine and classify its stationary points.

$$f'(x) = 3x^2 - 12x + 9, \qquad f'(x) = 0 \quad \Rightarrow \quad x = \{1, 3\}$$

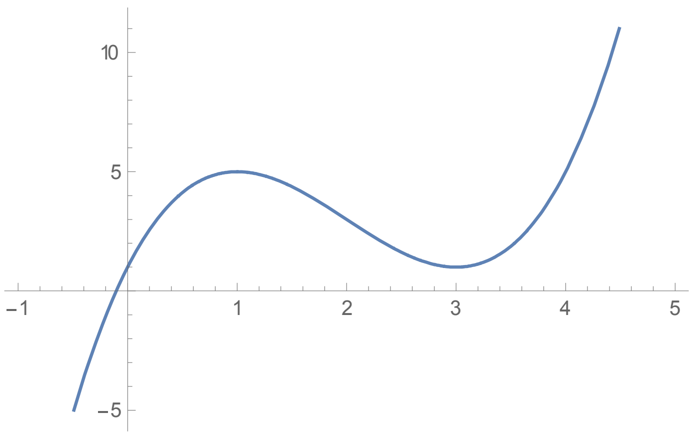
We can see that the derivative is zero at $x = 1$ and $x = 3$. The second derivative is

$$f''(x) = 6x - 12, \quad f''(1) = -6, \quad f''(3) = 6,$$

indicating that $x = 1$ is a maximum and $x = 3$ is a minimum. The sign of the first derivative is

$$\begin{cases} f'(x) > 0 & x \in (-\infty, 1) \\ f'(x) < 0 & x \in (1, 3) \\ f'(x) > 0 & x \in (3, \infty) \end{cases},$$

meaning that the function increases left of $x = 1$, decreases in the interval $(1, 3)$ and increases right of $x = 3$. The graph of the function, plotted by *Mathematica*, confirms these conclusions:



Multiple variable case

The multivariate case is richer in possibilities – stationary points can be maxima (Figure3a), minima (Figure 3b), but also *both* (Figure 3c): a point can be a minimum along one coordinate and a maximum along another. Such situations are called *saddle points*.

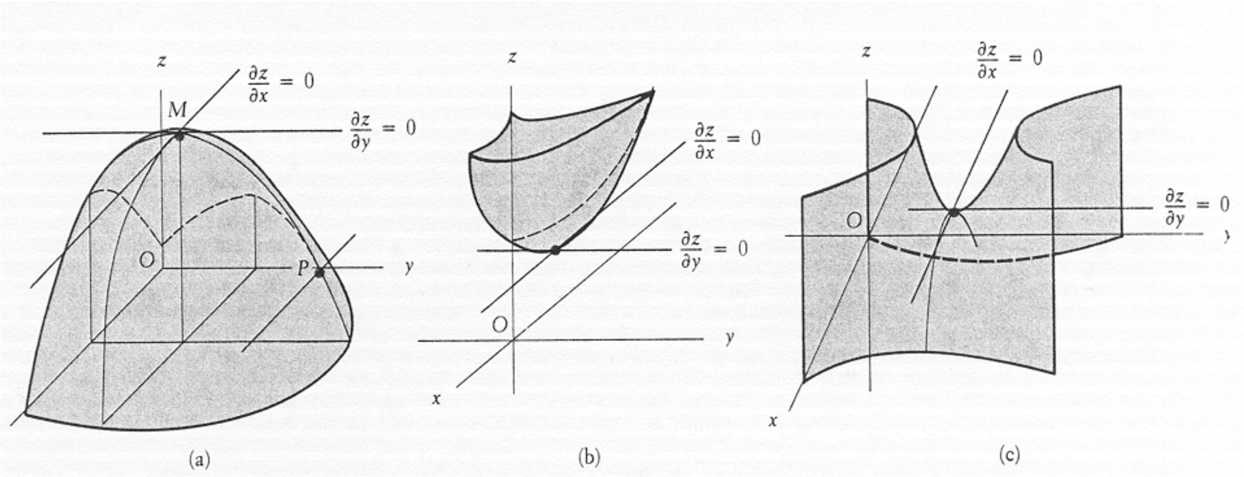


Figure 3. *Schematic drawings of a maximum (a), a minimum (b) and a saddle point for a function of two variables.*

The stationary point of a multivariate function must be a stationary point *in every direction*. This means that every partial derivative must be put to zero to obtain the list of stationary points of $f(x, y, ...)$:

$$\begin{cases} \partial f(x, y, ...)/\partial x = 0 \\ \partial f(x, y, ...)/\partial y = 0 \\ ... \end{cases} \tag{3}$$

The solutions of this system of equations are stationary points. Their classification is a complicated matter (we do not have sufficient time to cover it fully in this course), but for a function of two variables, solutions to the following quadratic equation with respect to λ determine the type:

$$\left(f''_{xx} - \lambda\right)\left(f''_{yy} - \lambda\right) - \left(f''_{xy}\right)^2 = 0 \tag{4}$$

Two positive solutions indicate a minimum, two negative solutions a maximum. One positive and one negative solution indicates a saddle point.

<u>Example 2</u>: by considering its partial derivatives with respect to both coordinates, find and classify the stationary points of $f(x, y) = (x-1)^2 - (y-2)^2$.

$$\frac{\partial f}{\partial x} = 2(x-1), \qquad \frac{\partial f}{\partial y} = -2(y-2)$$

Setting these derivatives to zero and solving the resulting system of equations yields:

$$\begin{cases} 2(x-1) = 0 \\ -2(y-2) = 0 \end{cases} \Rightarrow \begin{cases} x = 1 \\ y = 2 \end{cases}$$

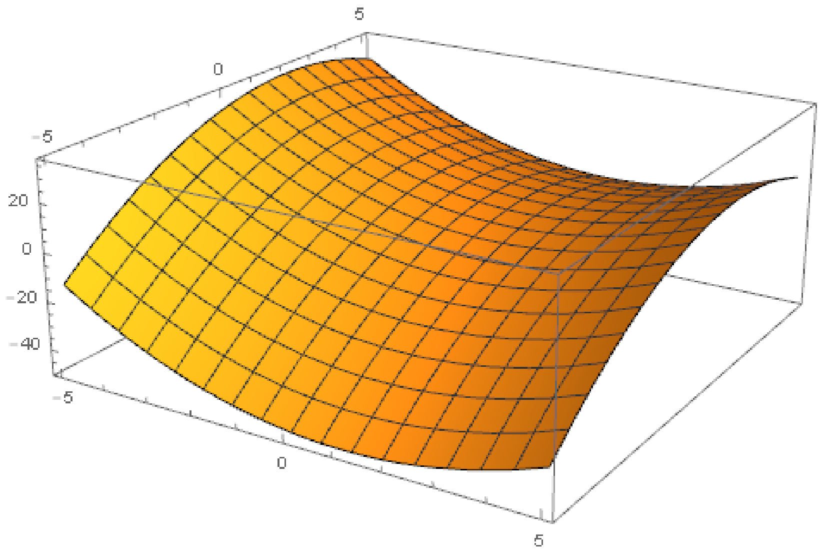
Computing the second derivatives at this point produces:

$$f''_{xx} = \frac{\partial^2 f}{\partial x^2} = 2, \qquad f''_{yy} = \frac{\partial^2 f}{\partial y^2} = -2, \qquad f''_{xy} = \frac{\partial^2 f}{\partial x \partial y} = 0$$

Plugging this into the diagnostic equation given above and solving it for λ yields:

$$\left(2-\lambda\right)\left(2+\lambda\right)=0 \quad \Rightarrow \quad \lambda=\pm 2$$

and therefore the stationary point in question is a saddle point. Plotting the function in *Mathematica* confirms this conclusion.



Numerical optimization of multivariate functions

In some cases the functions under consideration are too complicated for the system of equations involving partial derivatives to be soluble. In this case the location of stationary points is determined by a *search* procedure: a computer starts at some user-specified location and moves up or down the direction of steepest ascent. This direction is called the *gradient vector* $\vec{\nabla} f$:

$$\vec{\nabla} f\left(x, y, \ldots\right)=\begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} & \ldots \end{pmatrix}$$

It practice, it is computed by taking first partial derivatives and computing their values at a given point. Computer algorithms that climb or descend along the corresponding direction are called *gradient ascent* and *gradient descent* algorithms – they are implemented in all major computer algebra systems.

In the context of computational chemistry, the task of refining a molecular geometry from a given initial guess amounts to finding the minimum of the energy with respect to the atomic coordinates:

$$\left\{\vec{r}_1, \vec{r}_2, \ldots\right\}=\underset{\left\{\vec{r}_1, \vec{r}_2, \ldots\right\}}{\arg\min} E\left(\vec{r}_1, \vec{r}_2, \ldots\right) \tag{5}$$

where $\vec{r}_k=\left[x_k \quad y_k \quad z_k\right]$ are Cartesian coordinates of the k-th atom. The minimum of $E\left(\vec{r}_1, \vec{r}_2, \ldots\right)$ is called *equilibrium geometry*. Quantum chemistry software packages optimise molecular geometries using *gradient descent* methods, in which the geometry is updated by stepping down the energy gradient.

Week 30 workshop exercises

Steiner, 2nd edition: section 9.12, problems 21-26, 27(i), 28(i).

Demidovich, 2nd edition, problems 2008-2012.

Extra difficulty exercises for the brave

Demidovich, 2nd edition, problems 2013-2016.