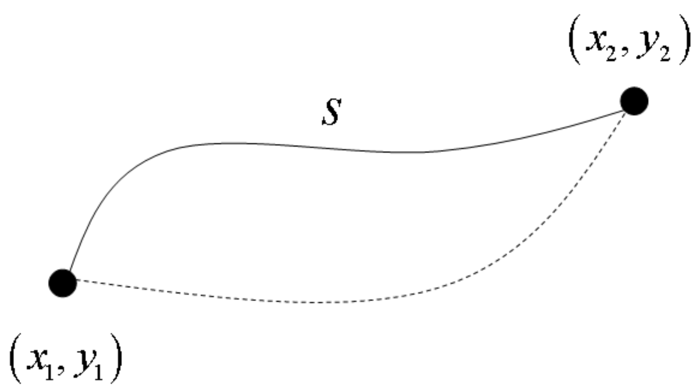
Minimisation of matrix-valued functions

Background:
Anybody with an elementary knowledge of calculus understands what we mean by a turning point of a function (a minimum, maximum or point of inflexion). For a 1-dimensional function $f(x)$, there is some value of the variable $x = x_{\min}$ within the allowed range which will result in the derivative of the function assuming a value of zero. The concept extends easily to multidimensional functions $f(x_1, x_2 ... x_n)$. In general, differential calculus is to a very large degree concerned with finding the minima and maxima of *known functions*.

In the calculus of variations, a more subtle concept is involved. Here we seek an *unknown function* that has the property that it will minimise or maximise a certain quantity which depends on the unknown function (such a quantity is technically known as a functional). Perhaps the simplest example of such a problem is to find the shortest path S between two points (x_1, y_1) and (x_2, y_2) in the 2-dimensional plane.



In this case, the total path length is the sum of all the infinitesimal elements *ds* from point 1 to point 2. We thus seek to minimise the integral –

$$Q = \int_{x_1,y_1}^{x_2,y_2} ds = \int_{x_1,y_1}^{x_2,y_2} \sqrt{dx^2 + dy^2} = \int_{x_1,y_1}^{x_2,y_2} \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx = \int_{x_1,y_1}^{x_2,y_2} \sqrt{1 + y'^2}\, dx = \int_{x_1,y_1}^{x_2,y_2} F(y, y'; x)\, dx$$

Thus the goal is to find some function $y(x)$ that minimises Q. Note that the functional F in this particular example depends explicitly only on the derivative of the function y' rather than the function itself. It is possible to show that the Euler-Lagrange equation can be used to solve this problem –

$$\frac{\partial F}{\partial x} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0$$

(Unsurprisingly, the answer is a straight line - as the reader may care to prove by a process of evaluating the Euler-Lagrange equation above and then integrating).

We will not repeat here the detailed argument which lies behind the derivation of the Euler-Lagrange equation. This can be found in many textbooks on the subjects and other sources. The key point we will make however is that we arrive at this equation by requiring that if the solution that minimises Q is a function $y_0(x)$, then the infinitesimal addition of some arbitrary function $\eta(x)$ so that $y_0(x) \to y_0(x) + \varepsilon\eta(x)$ where ε is small will result in a corresponding change in Q, $\delta Q = 0$. In other words, we require that the first variation in the quantity Q with respect to the function $y(x)$, denoted as δQ_y, vanishes.

Matrix variational notation

We want to extend the variational notion to matrices but first let's consider some preliminary issues.

To say that one number is bigger or smaller than another is quite unambiguous and clear in meaning. However, if we consider matrices which have many elements, the situation is not clear cut. What meaning might we assign to the notion that 'one matrix is bigger or smaller than another' ? It cannot simply be an issue of an element by element comparison because we will generally find that in comparing two matrices, some corresponding elements are bigger in one matrix and some in the other. The more meaningful question is to consider a matrix in terms of its deeper meaning – namely, a matrix effects a transformation on other matrices or vectors. Any combination of matrices is, in general, a matrix-valued function[1] and we want to find that value of a particular matrix at which the matrix-valued function is stationary.

Definitions:

i) Let **A** be a $n\,x\,n$ matrix and $\mathbf{x}$ be an $n\,x\,1$ vector then

$$Q = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

is called a quadratic form.

ii) If for all $\mathbf{x}$ except the null vector (all elements of $\mathbf{x}$ are zero) $Q = \mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ then **A** is said to be positive definite. If $Q = \mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$, **A** is said to be positive semi-definite.

iii) A positive definite matrix **A** is said to be smaller than a positive definite matrix **B** if **B-A** is positive definite.

iv) The positive definite matrix valued function *f* of the matrix **X** is said to possess a minimal if there exists an **X**, say $\mathbf{X}'$, such that $f(\mathbf{X}) - f(\mathbf{X}')$ **is positive semi-definite** for all **X** in an arbitrarily small neighbourhood of $\mathbf{X}'$

[1] They may produce a scalar (a number) in certain instances, of course.

Let the matrix function $f(\mathbf{X}) = \mathbf{XAX^T}$ be a positive definite matrix equation (i.e. $f(\mathbf{X})$ is matrix valued, not a scalar quantity). The first variation of $f(\mathbf{X})$ is produced by allowing a small variation in **X** so that $\mathbf{X} \to \mathbf{X} + \delta\mathbf{X}$ and thus $\delta f_X = f(\mathbf{X} + \delta\mathbf{X}) - f(\mathbf{X})$. **A condition for $f(\mathbf{X})$ to be minimal is that δf_X vanish as follows.** Consider that $f(\mathbf{X})$ is minimised for $\mathbf{X} = \mathbf{X}'$

$$\begin{aligned}\delta f_X &= f(\mathbf{X}' + \delta\mathbf{X}') - f(\mathbf{X}') \\ &= (\mathbf{X}' + \delta\mathbf{X}')\mathbf{A}(\mathbf{X}' + \delta\mathbf{X}')^{\mathbf{T}} - \mathbf{X}'\mathbf{A}\mathbf{X}'^{\mathbf{T}} \\ &= \delta\mathbf{X}'\mathbf{A}\mathbf{X}'^{\mathbf{T}} + \mathbf{X}'\mathbf{A}\delta\mathbf{X}'^{\mathbf{T}} + \delta\mathbf{X}'\mathbf{A}\delta\mathbf{X}'^{\mathbf{T}}\end{aligned}$$

Now $\delta\mathbf{X}'\mathbf{A}\delta\mathbf{X}'^{\mathbf{T}}$ is positive semi-definite and we conclude that $\mathbf{A}\mathbf{X}'^{\mathbf{T}} = 0$ or if **A** is non-singular (as is usually the case) that $\mathbf{X} = 0$.

Thus the variational concept is extended to matrices or matrix-valued functions by allowing perturbations or more correctly ***first variations*** in the matrix and requiring that these first variations vanish. In general, this will lead to a matrix equation for the minimising solution just as first variations in a function in the calculus of variations leads to a differential equation.