Optimization of a Functions under Constraints with the Lagrange[1] Multiplier Method

Optimization of an analytical function without constraints is at least in theory straightforward: [a] calculate the partial derivatives of the function with respect to its arguments; [b] set the derivatives equal to zero; [c] solve the system of partial derivative functions for the unknown values of the arguments, which may lead to more than one solution; and [d] check each solution whether it denotes a minimum, a maximum or a saddle point. The Lagrange multiplier method of finding optima under constraints is a logical extension of the unconstrained approach. It also provides useful analytical and geometrical interpretations of the constrained solution.

Motivating example: The objective function $f(x_1, x_2) = 14 + 4 \cdot x_1^2 - 4 \cdot x_1 \cdot x_2 + 4 \cdot x_2^2$ shown in Figure 1 clearly has one and only one minimum at the point $(x_1^* = 0, x_2^* = 0)$.

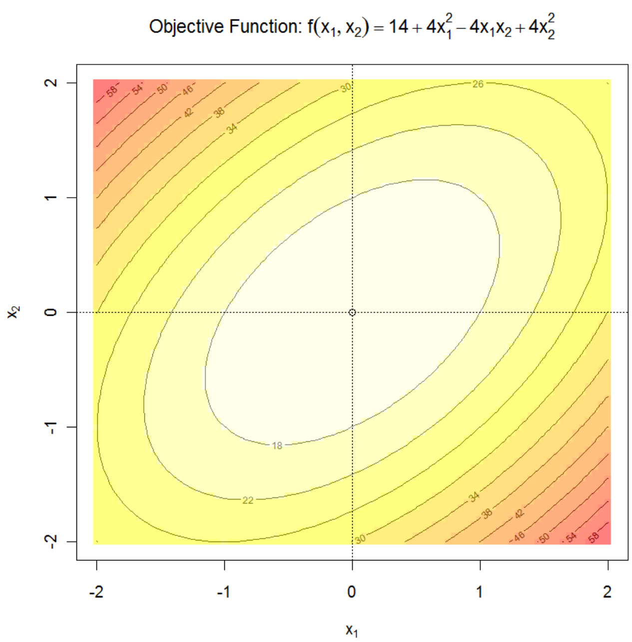

Figure 1: Evaluation of the objective function with the unconstrained optimum at (0,0)

This particular minimum may not be achieved if we restrict the search set for potential optimal points to just those that satisfy a condition such as $r(x_1, x_2) = 0.5 \cdot x_1 - x_2 \equiv c$, where the value c denotes one of many possible levels of the constraint. This particular constraint can be expressed as a linear function $x_2 = 0.5 \cdot x_1 - c$, and any solutions must lie on this line. Figure 2 plots this linear constraint at the given level $c \equiv 1$. For levels of $c > 1$ the constraint line moves in a parallel motion downward and for $c < 1$ it moves upwards.

---

[1] Joseph-Louis Lagrange (1736 –1813) was an Italian born mathematician and mathematical physicist who worked part of his life in Prussia and France. Other mathematicians of his time were Euler, who was also his “doctoral father”, and Gauss. Lagrange in turn was the “doctoral father” of Fourier.

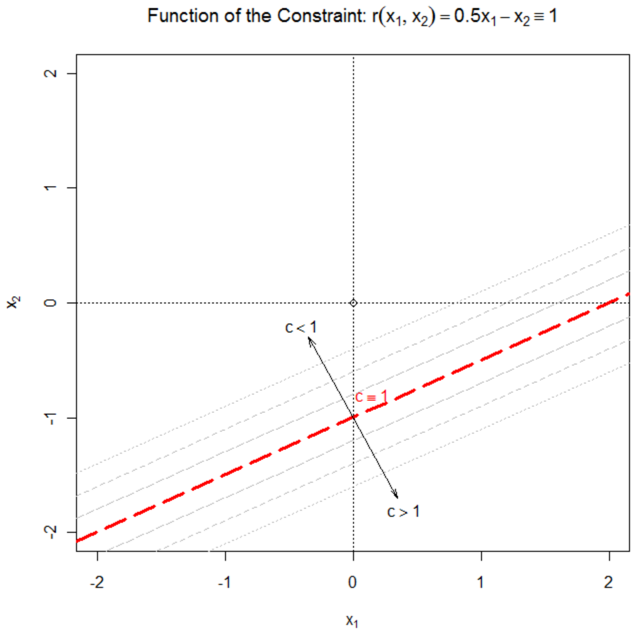

Figure 2: Simple linear constraint at the level $c \equiv 1$

General solution: We have to find an optimum of $f(x_1, x_2)$ within the set of points (x_1, x_2) that also are members of the constraint $r(x_1, x_2) \equiv c$. A solution to this constrained problem is obtained by the method of Lagrange multipliers. It proceeds by augmenting the objective function $f(x_1, x_2)$ with the constraint function, expressed in its equivalent zero form as $r(x_1, x_2) - c \equiv 0$, and attaches the multiplier λ to it. The augmented Lagrangian function to be optimized is:

$$\Lambda(x_1, x_2, \lambda) = f(x_1, x_2) - \lambda \cdot \left[r(x_1, x_2) - c \right].$$

Taking the partial derivatives of the augmented function $\Lambda(x_1, x_2, \lambda)$, setting them to zero and solving the system of equations for the augmented set of values $\{x_1, x_2, \lambda\}$ gives the necessary solution $(x_1^*, x_2^*, \lambda^*)$ (or a set solutions) after the appropriate validity checks of the sufficient condition (more on these later):

$$\begin{aligned}
\frac{\partial \Lambda(x_1, x_2, \lambda)}{\partial x_1} &= \frac{\partial f(x_1, x_2)}{\partial x_1} - \lambda \cdot \frac{\partial r(x_1, x_2)}{\partial x_1} \equiv 0 \\
\frac{\partial \Lambda(x_1, x_2, \lambda)}{\partial x_2} &= \frac{\partial f(x_1, x_2)}{\partial x_2} - \lambda \cdot \frac{\partial r(x_1, x_2)}{\partial x_2} \equiv 0 \\
\frac{\partial \Lambda(x_1, x_2, \lambda)}{\partial \lambda} &= -\left[r(x_1, x_2) - c \right] \equiv 0
\end{aligned}$$

The last partial derivative with respect to the multiplier λ is equivalent to the constraint at the level c. This derivative ensures that the solution will satisfy the constraint.

Analytically the multiplier λ^* can be interpreted as a *marginal change* in the constrained objective function after a relaxation of the constraint level c, that is, $\lambda^*(c) = \partial\Lambda^*(\ldots,c)/\partial c = \partial f\left(x_1^*(c), x_2^*(c)\right)/\partial c$. The values $x_1^*(c)$ and $x_2^*(c)$ denote the arguments that lead to an optimal value of the objective function $f(x_1, x_2)$ under the constraint $r(x_1, x_2) \equiv c$ and $\Lambda^*(\ldots,c)$ signifies that the augmented objective function at the solution implicitly depends on the level c. In other words, if we change the constraint *slightly* from c to $c + \Delta c$ then the value of the objective function changes. This rate of change can be approximated by

$$\frac{f\left(x_1^*(c), x_2^*(c)\right) - f\left(x_1^*(c+\Delta c), x_2^*(c+\Delta c)\right)}{\underbrace{c - (c + \Delta c)}_{=\Delta c}} \approx \lambda^* . \tag{1.1}$$

Extensions*[2]: The Lagrange multiplier method can be extended to incorporate more than two attributes, that is, $\{x_1, x_2, \ldots, x_k\}$, and more than one constraint $\{r_1(\,) \equiv c_1, r_2(\,) \equiv c_2, \ldots, r_q(\,) \equiv c_q\}$ including inequality constraints such as $r_p(x_1, x_2, \ldots, x_k) \leq c_p$ with $p \in \{1, 2, \ldots, q\}$. Note that constraints are not allowed to contradict each other because this would lead to an empty solution set.

Sufficient condition*: The evaluation of the set of possible solutions requires the calculation of the bordered Hessian matrix which is of dimension $[q+k] \times [q+k]$. This matrix comprises of all pair-wise second order partial derivatives of the *augmented* objective function $\partial^2\Lambda\left(x_1, x_2, \ldots, x_k\right)/\left(\partial x_i \cdot \partial x_j\right)$ and is bordered by blocks of $k \times q$ and $q \times k$, respectively, partial derivatives of the constraints $\partial r_p(x_1, x_2, \ldots, x_k)/\partial x_i$. A $q \times q$ block of zeros pads the upper left-hand corner. This bordered Hessian is evaluated at all possible solutions. If the sequence of minor determinants of increasing order alternate in their signs then that particular solution is a local or global maximum; if the sequence leads to all negative minor determinants then that particular solution is a local or global minimum.

For example, for $q = 1$ and $k = 2$ the bordered Hessian is

$$\underset{3\times 3}{\mathbf{H}} = \begin{pmatrix} 0 & \dfrac{\partial r(\,)}{\partial x_1} & \dfrac{\partial r(\,)}{\partial x_2} \\ \dfrac{\partial r(\,)}{\partial x_1} & \dfrac{\partial^2\Lambda(\,)}{\partial x_1 \cdot \partial x_1} & \dfrac{\partial^2\Lambda(\,)}{\partial x_1 \cdot \partial x_2} \\ \dfrac{\partial r(\,)}{\partial x_2} & \dfrac{\partial^2\Lambda(\,)}{\partial x_2 \cdot \partial x_1} & \dfrac{\partial^2\Lambda(\,)}{\partial x_2 \cdot \partial x_2} \end{pmatrix}.$$

Its only minor matrix is

[2] Topics labeled by a (*) may be skipped at the first reading.

$$\underset{2\times 2}{\mathbf{H}} = \begin{pmatrix} 0 & \dfrac{\partial r(\,)}{\partial x_1} \\ \dfrac{\partial r(\,)}{\partial x_1} & \dfrac{\partial^2 \Lambda(\,)}{\partial x_1 \cdot \partial x_1} \end{pmatrix}.$$

Both matrices need to be evaluated at the possible solutions $\left(x_1^*, x_2^*, \lambda^*\right)$. For a local or global maximum the determinants need to satisfy $\left|\underset{2\times 2}{\mathbf{H}}\left(x_1^*, x_2^*, \lambda^*\right)\right| < 0$ and $\left|\underset{3\times 3}{\mathbf{H}}\left(x_1^*, x_2^*, \lambda^*\right)\right| > 0$. For a local or global minimum both determinants must be negative.

<u>Continuation of the example:</u> The objective function is $f(x_1, x_2) = 14 + 4\cdot x_1^2 - 4\cdot x_1 \cdot x_2 + 4\cdot x_2^2$ and the constraint, after it has been expressed in the zero form, is $r(x_1, x_2) = 0.5\cdot x_1 - x_2 - c \equiv 0$ at an arbitrary level c:

1. The augmented objective function becomes

$$\Lambda(x_1, x_2, \lambda) = 14 + 4\cdot x_1^2 - 4\cdot x_1 \cdot x_2 + 4\cdot x_2^2 - \lambda \cdot \left(0.5\cdot x_1 - x_2 - c\right).$$

2. Its partial derivates, set to zero, are

$$\begin{aligned} \frac{\partial \Lambda(x_1, x_2, \lambda)}{\partial x_1} &= 8\cdot x_1 - 4\cdot x_2 - 0.5\cdot \lambda \equiv 0 \\ \frac{\partial \Lambda(x_1, x_2, \lambda)}{\partial x_2} &= -4\cdot x_1 + 8\cdot x_2 + \lambda \equiv 0 \\ \frac{\partial \Lambda(x_1, x_2, \lambda)}{\partial \lambda} &= -0.5\cdot x_1 + x_2 + c \equiv 0 \end{aligned} \quad . \qquad (1.2)$$

Note that the last derivative equals to the constraint.

3. This is a system of partial derivatives are linear functions in the arguments and can be arranged in matrix terms

$$\begin{pmatrix} 8 & -4 & -0.5 \\ -4 & 8 & 1 \\ -0.5 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1^*(c) \\ x_2^*(c) \\ \lambda^*(c) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -c \end{pmatrix}.$$

Setting the level of the constraint to $c \equiv 1$ gives

$$\begin{pmatrix} 0 \\ -1 \\ 8 \end{pmatrix} = \begin{pmatrix} 8 & -4 & -0.5 \\ -4 & 8 & 1 \\ -0.5 & 1 & 0 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}. \qquad (1.3)$$

Consequently as shown in Figure 3, at the solution $x_1^*(c \equiv 1) = 0$, $x_2^*(c \equiv 1) = -1$ the constrained minimum $f(0,-1) = 18$ is attained, which clearly differs from the unconstrained minimum. As required this solution also satisfies the constraint $r(0,-1) = 0.5 \cdot 0 + 1 \equiv 1$. The solution for the Lagrange multiplier is $\lambda^*(c \equiv 1) = 8$.

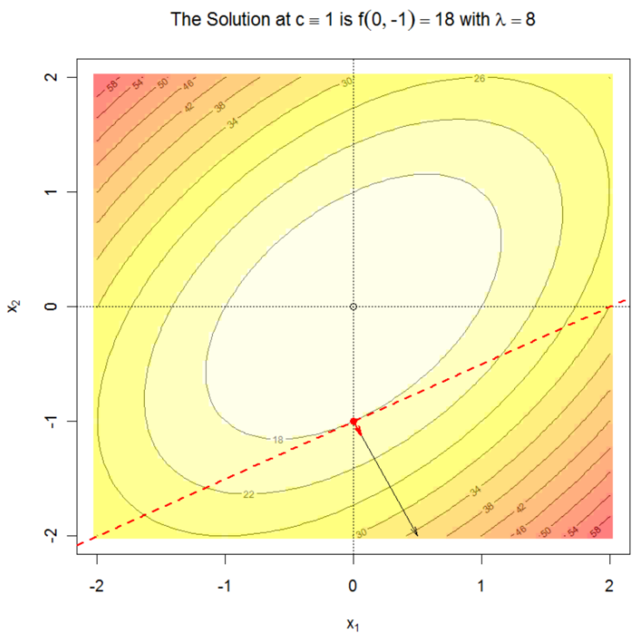

Figure 3: Solution of the constrained optimization problem

Geometric properties of the solution: Figure 3 shows a general property of the Lagrange multiplier solution. Its constraint function tangentially touches the isoline[3] of the objective function at the restricted optimum but it does not cross it. Would it cross the isoline then there would be another isoline with an improved optimum. The gradient vectors $\nabla_{x_1,x_2} f(x_1,x_2) = \left(\frac{\partial f(x_1,x_2)}{\partial x_1}, \frac{\partial f(x_1,x_2)}{\partial x_2} \right)^T$ of the objective function are always perpendicular to isolines and the gradient vectors $\nabla_{x_1,x_2} r(x_1,x_2) = \left(\frac{\partial r(x_1,x_2)}{\partial x_1}, \frac{\partial r(x_1,x_2)}{\partial x_2} \right)^T$ of the constraint function are always perpendicular to the constraint. The gradients vectors can be evaluated at any point (x_1, x_2) and they express for the objective function the direction of the fastest ascend. Both gradients must be parallel at the tangential point of the optimal solution, that is, they point either into the same direction or into opposite directions. This property can be expressed by the equality

$$\nabla_{x_1,x_2} f(x_1^*, x_2^*) = \lambda^* \cdot \nabla_{x_1,x_2} r(x_1^*, x_2^*), \quad (1.4)$$

[3] In this case an isoline denotes that set of points (x_1, x_2) for which the objective function remains constant, i.e., $f(x_1, x_2) = \text{constant}$

where the scalar Lagrange multiplier λ^* ensures the proportionality of both gradients. This property provides a geometric interpretation of the Lagrange multiplier λ^*. The gradient of the objective function is $\nabla_{x_1,x_2} f(x_1^*, x_2^*) = (0,0)^T$ at its *unconstrained* optimum, that is, all partial derivates are zero. Recall that this property is exploited to find unconstrained optimum. Should the constrained and the unconstrained optimal solutions happen to be coincide then the Lagrange multiplier must be $\lambda^* = 0$, because under general conditions $\nabla_{x_1,x_2} r(x_1^*, x_2^*) \neq (0,0)^T$.

<u>Continuation of the example:</u> From equations (1.2) and the solution at (1.3) we obtain the gradients $\nabla_{x_1,x_2} f(x_1^*, x_2^*) = (4,-8)^T$ and $\nabla_{x_1,x_2} r(x_1^*, x_2^*) = (0.5,-1)^T$. They jointly satisfy equation (1.4) for the multiplier $\lambda^* = 8$. Figure 3 shows both gradient vectors after they have been jointly rescaled by the factor $\frac{1}{8}$ for display purposes. Note that because the constraint is a linear function, its gradient is constant at any level c. This will not hold for nonlinear constraints. Also a level $c = 0$ of the constraint does not imply under general conditions that the constrained and unconstrained optima are identical.

<u>Try it yourself:</u> You can find the R-script for this example in the file **`Lagrange01.r`**. If you change the constraint value **`c.val`** in line 5, you can observe how the constraint shifts and subsequently how the value of the objective function changes. Using equation (1.1) the rate of change in the objective function can be evaluated. For tiny changes it should be close to the Lagrange multiplier $\lambda(c)$. The script also demonstrates how to obtain basic symbolic derivatives in with the function **`deriv()`** and how to label graphs with mathematical expression by using the **`expression()`** function or the **`bquote()`** function in combination with the **`.()`** operator. See syntax details in the **`plotmath`** online.

<u>An elaborated example[*]:</u> Slightly more complicated objective and constraint functions lead to a set of first order partial derivatives that can no longer be expressed as a system of linear functions. Then numerical solutions strategies need to be sought to find $(x_1^*, x_2^*, \lambda^*)$ at a given constraint level c. An example provide the objective function $f(x_1, x_2) = 14 + 3 \cdot x_1^2 - 4 \cdot x_1 \cdot x_2 + 4 \cdot x_2^2 + \frac{1}{2} \cdot x_1 - 2 \cdot x_2$ and the constraint function $r(x_1, x_2) = \frac{1}{2} \cdot x_1 + x_1^2 - x_2 \equiv c$. The script **`Lagrange02.r`** shows how to address the problem of finding a solution for a set of non-linear functions.

Note that solutions of the non-linear equation solver **`nleqslv()`** (line 26) are sensitive to the starting values. Figure 4 below shows that the correct results for the *global minimum* were obtained for a starting value $\left\{x_1^{start} = 1, x_2^{start} = 1, \lambda^{start} = 0\right\}$. The dashed blue line in that figure outlines the correct path of the constrained global minima for varying levels of c. In contrast, Figure 5 demonstrates that the function **`nleqslv()`** provides unwarranted global minimum solutions for improperly chosen starting values $\left\{x_1^{start} = 0, x_2^{start} = 0, \lambda^{start} = 0\right\}$ and constraint levels[4] $c > 0$. In fact, as an evaluation of the determinant criterion for the bordered Hessian shows (see the script **`Lagrange02.r`**), the suggested solutions are then *local maxima*. This is not a fault of the function **`nleqslv()`** *per se*: non-linear

[4] This level is not exact and varies in dependence of the starting value in close proximity to $c \simeq 0$.

systems of equations can have more than one solution. So depending on the starting value, the non-linear equation solver just converges to a different solution. Numerically that solution is not incorrect because equation (1.4) still holds and both gradient vectors are parallel. One could get yet even another unwarranted but numerically correct solutions by setting the starting value to $\{x_1^{start} = -1, x_2^{start} = 1, \lambda^{start} = 0\}$ and selecting $c > 0$. This leads to *local minima* (see Figure 6). Irrespectively of the starting value, for constraint levels of $c < 0$ the solution path does not depend on the starting value and there is only one unique solution, which is the *global minimum* (see Figure 7).

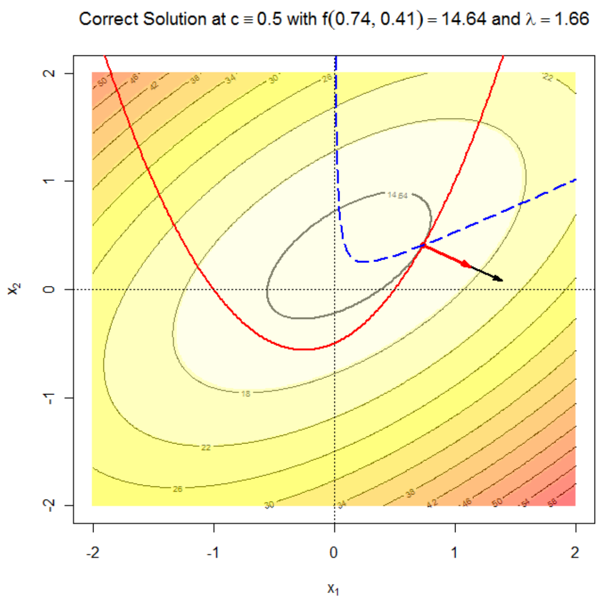

Figure 4: Correct global solution with a *global minimum* at starting value (1,1,0)

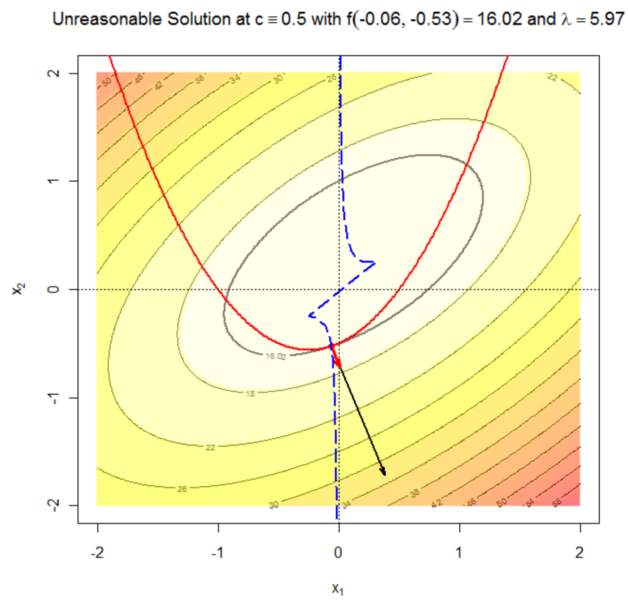

Figure 5: Unreasonable solution with a *local maximum* at starting value (0,0,0)

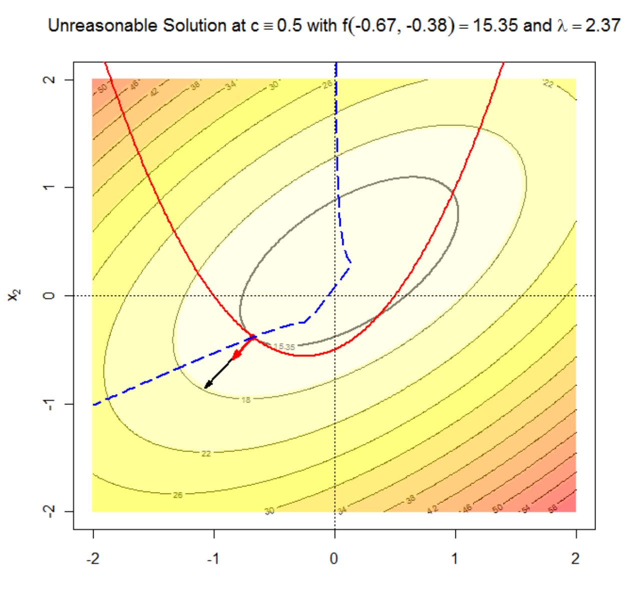

Figure 6: Unreasonable solution with a *local minimum* at starting value (-1,1,0)

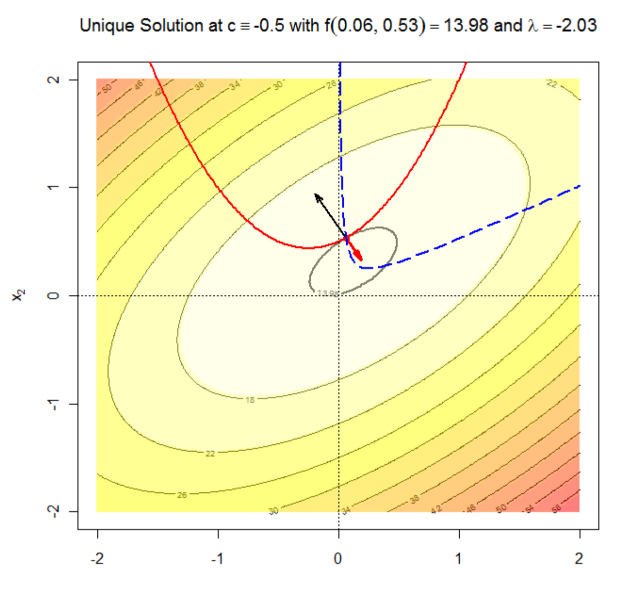

Figure 7: For c<0 there is only one unique solutions, which is the *global minimum*

General lesson to be learned here: ***Be always critical and check your solution approach over a reasonable range of feasible inputs***. This applies not only to the non-linear optimization example, which under particular conditions suggests solutions other than the appropriate one for the global minimum and which depends very much on the initialization of the algorithm to solve a non-linear system of equations. Consequently, expose your GIS models and analyses to the possibility of failure. Review your intermediate analyses and results and perform validity checks. ***Formulate what you expect to find and then compare your expectations with the suggested results.*** Should the results not match your expectations then either you have messed up in the preceding steps or the software implementation is faulty. Or – after the possibility of errors has been ruled out – you must revise your expectations and perhaps the theory underlying them. This process of revision is one instance where academic progress is made and the body of knowledge is enhanced!