4. Newton Search for a Minimum

Newton's Method

The quadratic approximation method for finding a minimum of a function of one variable generated a sequence of second degree Lagrange polynomials, and used them to approximate where the minimum is located. It was implicitly assumed that near the minimum, the shape of the quadratics approximated the shape of the objective function $y = f(x)$. The resulting sequence of minimums of the quadratics produced a sequence converging to the minimum of the objective function $y = f(x)$. Newton's search method extends this process to functions of n independent variables: $f(\vec{\mathbf{X}}) = f(x_1, x_2, \ldots, x_n)$. Starting at an initial point $\vec{\mathbf{P}}_0$, a sequence of second-degree polynomials in n variables can be constructed recursively. If the objective function is well-behaved and the initial point is near the actual minimum, then the sequence of minimums of the quadratics will converge to the minimum of the objective function. The process will use both the first- and second-order partial derivatives of the objective function. Recall that the gradient method used only the first partial derivatives. It is to be expected that Newton's method will be more efficient than the gradient method.

Background

Now we turn to the minimization of a function $f(\vec{\mathbf{X}})$ of n variables, where $\vec{\mathbf{X}} = (x_1, x_2, \ldots, x_n)$ and the partial derivatives of $f(\vec{\mathbf{X}})$ are accessible. Although the Newton search method will turn out to have a familiar form. For illustration purposes we emphasize the two dimensional case when $f(\vec{\mathbf{X}}) = f(x, y)$. The extension to n dimensions is discussed in the hyperlink.

Definition (Gradient). Assume that $f(\vec{\mathbf{X}}) = f(x, y)$ is a function of two variables, $\vec{\mathbf{X}} = (x, y)$, and has partial derivatives $\frac{\partial f(x,y)}{\partial x}$ and $\frac{\partial f(x,y)}{\partial x}$. The gradient of $f(\vec{\mathbf{X}})$, denoted by $\nabla f(\vec{\mathbf{X}})$, is the vector function

$$\nabla f(\vec{\mathbf{X}}) = \nabla f(x, y) = \left(\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y} \right) = (f_x(x, y), f_y(x, y)).$$

Definition (Jacobian Matrix). Assume that $f_1(x, y)$ and $f_2(x, y)$ are functions of two variables, $\vec{\mathbf{X}} = (x, y)$, their Jacobian matrix $\mathbf{J}(x, y) = \mathbf{J}[f_1, f_2]$ is

$$\mathbf{J}(\vec{\mathbf{X}}) = \mathbf{J}(x, y) = \begin{pmatrix} \frac{\partial f_1(x,y)}{\partial x} & \frac{\partial f_1(x,y)}{\partial y} \\ \frac{\partial f_2(x,y)}{\partial x} & \frac{\partial f_2(x,y)}{\partial y} \end{pmatrix}.$$

Definition (Hessian Matrix). Assume that $f(\vec{\mathbf{X}}) = f(x, y)$ is a function of two variables, $\vec{\mathbf{X}} = (x, y)$, and has partial derivatives up to the order two. The Hessian matrix $\mathbf{H}(x, y)$ is defined as follows:

$$\mathbf{H}(\vec{\mathbf{X}}) = \mathbf{H}(x, y) = \begin{pmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{pmatrix}.$$

Lemma 1. For $f(x, y)$ the Hessian matrix $\mathbf{H}(x, y)$ is the Jacobian matrix for the two functions $f_x(x, y)$ and $f_y(x, y)$, i. e.

$$\mathbf{H}(x, y) = \begin{pmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_x(x,y)}{\partial x} & \frac{\partial f_x(x,y)}{\partial y} \\ \frac{\partial f_y(x,y)}{\partial x} & \frac{\partial f_y(x,y)}{\partial y} \end{pmatrix} = \mathbf{J}(x, y).$$

Lemma 2. If the second order partial derivatives of $f(\vec{\mathbf{X}})$ are continuous then the Hessian matrix $\mathbf{H}(\vec{\mathbf{X}})$ is symmetric.

Example 1. Find the gradient vector and Hessian matrix at the point $(-3, -2)$ for the function $f(x, y) = \dfrac{x - y}{x^2 + y^2 + 2}$.

Solution 1.

Taylor Polynomial for f(x,y)

Assume that $f(\vec{\mathbf{X}}) = f(x, y)$ is a function of two variables, $\vec{\mathbf{X}} = (x, y)$, and has partial derivatives up to the order two. There are two ways to write the quadratic approximation to f(x,y), based on series and matrices, respectfully. Assume that the point of expansion is $\vec{\mathbf{P}} = (p, q)$ and use the notation $\mathbf{\Delta}\vec{\mathbf{P}} = (\Delta p, \Delta q)$ and $\vec{\mathbf{X}} = \vec{\mathbf{P}} + \mathbf{\Delta}\vec{\mathbf{P}}$, then

$$\vec{\mathbf{X}} = (x, y) = (p + \Delta p, q + \Delta q) = (p, q) + (\Delta p, \Delta q) = \vec{\mathbf{P}} + \mathbf{\Delta}\vec{\mathbf{P}}.$$

The Taylor polynomial using series notation is

$$f(p + \Delta p, q + \Delta q) \approx f(p, q) + f_x(p, q)\,\Delta p + f_y(p, q)\,\Delta q + \frac{1}{2!}\left(f_{xx}(p, q)\,\Delta p^2 + 2 f_{xy}(p, q)\,\Delta p\,\Delta q + f_{yy}(p, q)\,\Delta q^2\right)$$

The Taylor polynomial using vector and matrix notation is

$$f(p + \Delta p, q + \Delta q) \approx f(p, q) + \{f_x(p, q), f_y(p, q)\}.\begin{pmatrix} \Delta p \\ \Delta q \end{pmatrix} + \frac{1}{2!}\{\Delta p, \Delta q\}.\begin{pmatrix} f_{xx}(p, q) & f_{xy}(p, q) \\ f_{yx}(p, q) & f_{yy}(p, q) \end{pmatrix}.\begin{pmatrix} \Delta p \\ \Delta q \end{pmatrix}$$

The latter can be written in the form

$$f(p+\Delta p, q+\Delta q) \approx f(p,q) + \nabla f(p,q).\begin{pmatrix}\Delta p\\ \Delta q\end{pmatrix} + \frac{1}{2!}\{\Delta p, \Delta q\}.\mathbf{H}(p,q).\begin{pmatrix}\Delta p\\ \Delta q\end{pmatrix}.$$

Using vector notations $\vec{\mathbf{P}}$, $\Delta\vec{\mathbf{P}}$ and $\vec{\mathbf{X}} = \vec{\mathbf{P}} + \Delta\vec{\mathbf{P}}$ it looks like

$$f(\vec{\mathbf{X}}) \approx f(\vec{\mathbf{P}}) + \nabla f(\vec{\mathbf{P}}).\left(\Delta\vec{\mathbf{P}}\right) + \frac{1}{2!}\left(\Delta\vec{\mathbf{P}}\right).\mathbf{H}(\vec{\mathbf{P}}).\left(\Delta\vec{\mathbf{P}}\right)^{T}.$$

The above formula is also the expansion of $f(\vec{\mathbf{X}})$ centered at the point $\vec{\mathbf{P}}$ with $\Delta\vec{\mathbf{P}} = \vec{\mathbf{X}} - \vec{\mathbf{P}}$.

Example 2. Calculate the second-degree Taylor polynomial of $f(x, y) = \dfrac{x-y}{x^2+y^2+2}$ centered at the point $\vec{\mathbf{P}} = (-3, -2)$.
Solution 2.

Newton's Method for Finding a Minimum

Now we turn to the minimization of a function $f(\vec{\mathbf{X}})$ of n variables, where $\vec{\mathbf{X}} = (x_1, x_2, \ldots, x_n)$ and the partial derivatives of $f(\vec{\mathbf{X}})$ are accessible. Assume that the first and second partial derivatives of $w = f(\vec{\mathbf{X}}) = f(x_1, x_2, \ldots, x_n)$ exist and are continuous in a region containing the point $\vec{\mathbf{P}}_0$, and that there is a minimum at the point $\vec{\mathbf{P}}$. The quadratic polynomial approximation to $f(\vec{\mathbf{X}})$ is

$$Q(\vec{\mathbf{X}}) = f(\vec{\mathbf{P}}_0) + \nabla f(\vec{\mathbf{P}}_0).\left(\Delta\vec{\mathbf{P}}_0\right) + \frac{1}{2!}\left(\Delta\vec{\mathbf{P}}_0\right).\mathbf{H}(\vec{\mathbf{P}}_0).\left(\Delta\vec{\mathbf{P}}_0\right)^{T}$$

A minimum of $Q(\vec{\mathbf{X}})$ occurs where $\nabla Q(\vec{\mathbf{X}}) = \vec{\mathbf{0}}$.

Using the notation $\vec{\mathbf{P}}_0 = (p_0, q_0)$ and $\Delta\vec{\mathbf{P}}_0 = \vec{\mathbf{X}} - \vec{\mathbf{P}}_0 = (x-p_0, y-q_0)$ and the symmetry of $\mathbf{H}(\vec{\mathbf{P}}_0)$, we write

$$\begin{aligned}Q(x,y) = {} & f(p_0,q_0) + \{f_x(p_0,q_0), f_y(p_0,q_0)\}.\begin{pmatrix}x-p_0\\ y-q_0\end{pmatrix}\\ & + \frac{1}{2!}\{x-p_0, y-q_0\}.\begin{pmatrix}f_{xx}(p_0,q_0) & f_{xy}(p_0,q_0)\\ f_{xy}(p_0,q_0) & f_{yy}(p_0,q_0)\end{pmatrix}.\begin{pmatrix}x-p_0\\ y-q_0\end{pmatrix}\end{aligned}$$

The gradient $\nabla Q(\vec{\mathbf{X}})$ is given by the calculation

$$\nabla Q(x,y) = \{f_x(p_0,q_0), f_y(p_0,q_0)\} + \{x-p_0, y-q_0\}.\begin{pmatrix}f_{xx}(p_0,q_0) & f_{xy}(p_0,q_0)\\ f_{xy}(p_0,q_0) & f_{yy}(p_0,q_0)\end{pmatrix}$$

Thus the expression for $\nabla Q(\vec{\mathbf{X}}) = \vec{\mathbf{0}}$ can be written as

$$\nabla f(\vec{P}_0) + (\vec{X} - \vec{P}_0) . H(\vec{P}_0) = \vec{0}.$$

If $\vec{P}_0$ is close to the point $\vec{P}$ (where a minimum of f occurs), then $H(\vec{P}_0)$ is invertible the above equation can be solved for $\vec{X}$, and we have

$$\vec{X} = \vec{P}_0 - \nabla f(\vec{P}_0) \left(H(\vec{P}_0)\right)^{-1}.$$

This value of $\vec{X}$ can be used as the next approximation to $\vec{P}$ and is the first step in Newton's method for finding a minimum

$$\vec{P}_1 = \vec{P}_0 - \nabla f(\vec{P}_0) \left(H(\vec{P}_0)\right)^{-1}.$$

Lemma If column vectors are preferred over row vectors, then $(\vec{P}_1)^T$ is given by the computation

$$(\vec{P}_1)^T = (\vec{P}_0)^T - \left(H(\vec{P}_0)\right)^{-1} \left(\nabla f(\vec{P}_0)\right)^T.$$

Remark. This is equivalent to a Newton-Raphson root finding problem: Given the vector function $\vec{F}(\tilde{X}) = \left(\nabla f(\vec{X})\right)^T$ find the root of the equation $\vec{F}(\tilde{X}) = \left(\nabla f(\vec{X})\right)^T = \tilde{0}$. For this problem the Newton-Raphson formula is known to be

$$\tilde{P}_1 = \tilde{P}_0 - \left(J(\tilde{P}_0)\right)^{-1} \vec{F}(\tilde{P}_0),$$

where our previous work with Newton-Raphson method used column vectors $\tilde{P}_1$ and $\tilde{P}_0$. Here we use $\vec{F}(\tilde{P}_0) = \left(\nabla f(\vec{P}_0)\right)^T$ and Lemma 1 gives the relationship $\left(J(\tilde{P}_0)\right)^{-1} = \left(H(\vec{P}_0)\right)^{-1}$.

Outline of the Newton Method for Finding a Minimum

Start with the approximation $\vec{P}_0$ to the minimum point $\vec{P}$. Set $k = 0$.

(i) Evaluate the gradient vector $\nabla f(\vec{P}_k)$ and Hessian matrix $H(\vec{P}_k)$

(ii) Compute the next point $\vec{P}_{k+1} = \vec{P}_k - \nabla f(\vec{P}_k) \left(H(\vec{P}_k)\right)^{-1}$.

(iii) Perform the termination test for minimization. Set $k = k + 1$.

Repeat the process.

In equation (ii) the inverse of the Hessian matrix is used to solve for $\vec{P}_{k+1}$. It would be better to solve the system of linear equations represented by equation (ii) with a linear system solver rather than a matrix inversion. The reader should realize that the inverse is primarily a theoretical tool and the computation and use of inverses is inherently inefficient.

Example 3. Use the Newton search method to find the minimum of
$f(x, y) = x^2 - 4x + y^2 - y - xy.$
Solution 3.

Example 4. Use the Newton method to find the minimum of
$f(x, y) = 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4 + \frac{1}{4}x.$
Looking at your graphs, estimate the location of the local minima.
Solution 4.

Example 1. Find the gradient vector and Hessian matrix at the point $(-3, -2)$ for the function $f(x, y) = \dfrac{x - y}{x^2 + y^2 + 2}$.

Solution 1.

Enter the function, compute the first order partial derivatives and compute the gradient.

$$f[x,y] = \frac{x - y}{2 + x^2 + y^2}$$

$$f_x[x,y] = \frac{2 - x^2 + 2xy + y^2}{(2 + x^2 + y^2)^2}$$

$$f_y[x,y] = \frac{-2 - x^2 - 2xy + y^2}{(2 + x^2 + y^2)^2}$$

The gradient is

$$\nabla f[x,y] = \{f_x[x,y], f_y[x,y]\}$$

$$\nabla f[x,y] = \left\{\frac{2 - x^2 + 2xy + y^2}{(2 + x^2 + y^2)^2}, \frac{-2 - x^2 - 2xy + y^2}{(2 + x^2 + y^2)^2}\right\}$$

The gradient at the point $(-3, -2)$ is

$$\nabla f[-3,-2] = \left\{\frac{1}{25}, -\frac{19}{225}\right\}$$

Compute the second order partial derivatives and compute the Hessian matrix.

$$f[x,y] = \frac{x - y}{2 + x^2 + y^2}$$

$$f_{xx}[x,y] = \frac{2\,(-6\,x + x^3 + 2\,y - 3\,x^2\,y - 3\,x\,y^2 + y^3)}{(2 + x^2 + y^2)^3}$$

$$f_{xy}[x,y] = \frac{2\,(2\,x + x^3 - 2\,y + 3\,x^2\,y - 3\,x\,y^2 - y^3)}{(2 + x^2 + y^2)^3}$$

$$f_{yy}[x,y] = -\frac{2\,(2\,x + x^3 - 6\,y - 3\,x^2\,y - 3\,x\,y^2 + y^3)}{(2 + x^2 + y^2)^3}$$

The Hessian matrix is

$$H[x,y] = \begin{pmatrix} f_{xx}[x,y] & f_{xy}[x,y] \\ f_{xy}[x,y] & f_{yy}[x,y] \end{pmatrix}$$

$$H[x,y] = \begin{pmatrix} \frac{2\,(-6\,x+x^3+2\,y-3\,x^2\,y-3\,x\,y^2+y^3)}{(2+x^2+y^2)^3} & \frac{2\,(2\,x+x^3-2\,y+3\,x^2\,y-3\,x\,y^2-y^3)}{(2+x^2+y^2)^3} \\ \frac{2\,(2\,x+x^3-2\,y+3\,x^2\,y-3\,x\,y^2-y^3)}{(2+x^2+y^2)^3} & -\frac{2\,(2\,x+x^3-6\,y-3\,x^2\,y-3\,x\,y^2+y^3)}{(2+x^2+y^2)^3} \end{pmatrix}$$

The Hessian matrix at the point (-3, -2) is

$$H[-3,-2] = \begin{pmatrix} \frac{46}{1125} & -\frac{26}{1125} \\ -\frac{26}{1125} & -\frac{122}{3375} \end{pmatrix}$$

Example 2. Calculate the second-degree Taylor polynomial of $f(x, y) = \dfrac{x - y}{x^2 + y^2 + 2}$ centered at the point $\vec{\mathbf{P}} = (-3, -2)$.

Solution 2.

$$\vec{X} = \{x, y\}$$

$$f[\vec{X}] = \frac{x - y}{2 + x^2 + y^2}$$

$$\vec{P} = \{-3, -2\}$$

$$\overrightarrow{\Delta P} = \vec{X} - \vec{P}$$

$$\overrightarrow{\Delta P} = \{x, y\} - \{-3, -2\} = \{3 + x, 2 + y\}$$

$$f[\vec{P}] = -\frac{1}{15}$$

$$\nabla f[\vec{P}] = \left\{\frac{1}{25}, -\frac{19}{225}\right\}$$

$$H[\vec{P}] = \begin{pmatrix} \frac{46}{1125} & -\frac{26}{1125} \\ -\frac{26}{1125} & -\frac{122}{3375} \end{pmatrix}$$

$$f[\{x, y\}] = f[\vec{P}] + \text{gradF}[\vec{P}].\overrightarrow{\Delta P} + \frac{1}{2!}(\overrightarrow{\Delta P}).H[\vec{P}].(\overrightarrow{\Delta P})$$

$$Q[\{x,y\}] = -\frac{1}{15} + \left\{\frac{1}{25}, -\frac{19}{225}\right\}\{3 + x, 2 + y\} + \frac{1}{2}\{3 + x, 2 + y\}\begin{pmatrix} \frac{46}{1125} & -\frac{26}{1125} \\ -\frac{26}{1125} & -\frac{122}{3375} \end{pmatrix}\{3 + x, 2 + y\}$$

$$Q[\{x,y\}] = -\frac{1}{15} + \frac{3 + x}{25} - \frac{19(2 + y)}{225} + \frac{1}{2}\{3 + x, 2 + y\}\begin{pmatrix} \frac{46}{1125} & -\frac{26}{1125} \\ -\frac{26}{1125} & -\frac{122}{3375} \end{pmatrix}\{3 + x, 2 + y\}$$

$$Q[\{x,y\}] = -\frac{1}{15} + \frac{3 + x}{25} - \frac{19(2 + y)}{225} + \frac{1}{2}\{3 + x, 2 + y\}\begin{pmatrix} \frac{46(3+x)}{1125} - \frac{26(2+y)}{1125} \\ -\frac{26(3+x)}{1125} - \frac{122(2+y)}{3375} \end{pmatrix}$$

$$Q[\{x,y\}] = -\frac{1}{15} + \frac{3 + x}{25} - \frac{19(2 + y)}{225} + \frac{1}{2}\left(\frac{46(3 + x)^2}{1125} - \frac{52(3 + x)(2 + y)}{1125} - \frac{122(2 + y)^2}{3375}\right)$$

The second-degree Taylor polynomial is

$$Q[\{x,y\}] = -\frac{1}{15} + \frac{3+x}{25} - \frac{19\,(2+y)}{225} + \frac{23\,(3+x)^2}{1125} - \frac{26\,(3+x)\,(2+y)}{1125} - \frac{61\,(2+y)^2}{3375}$$

Which can be simplified

$$Q[\{x,y\}] = -\frac{1}{15} + -\frac{11}{225} + \frac{x}{25} - \frac{19\,y}{225} + \left(-\frac{91}{3375} + \frac{86\,x}{1125} + \frac{23\,x^2}{1125} - \frac{478\,y}{3375} - \frac{26\,x\,y}{1125} - \frac{61\,y^2}{3375}\right)$$

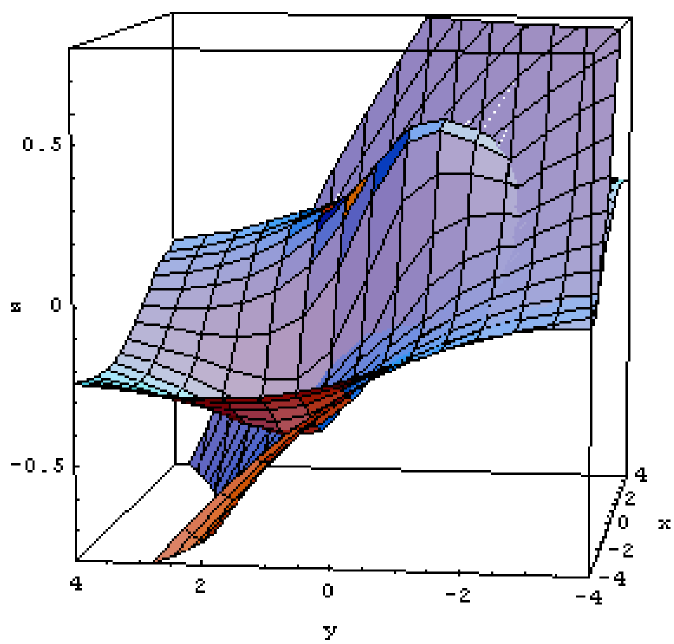
$$Q[\{x,y\}] = -\frac{1}{15} + \frac{3+x}{25} - \frac{19\,(2+y)}{225} + \frac{23\,(3+x)^2}{1125} - \frac{26\,(3+x)\,(2+y)}{1125} - \frac{61\,(2+y)^2}{3375}$$

$$Q[\{x,y\}] = -\frac{481}{3375} + \frac{131\,x}{1125} + \frac{23\,x^2}{1125} - \frac{763\,y}{3375} - \frac{26\,x\,y}{1125} - \frac{61\,y^2}{3375}$$

In simplified form we have

$$Q[\{x,y\}] = \frac{-481 + 393\,x + 69\,x^2 - 763\,y - 78\,x\,y - 61\,y^2}{3375}$$

We can view the surfaces.

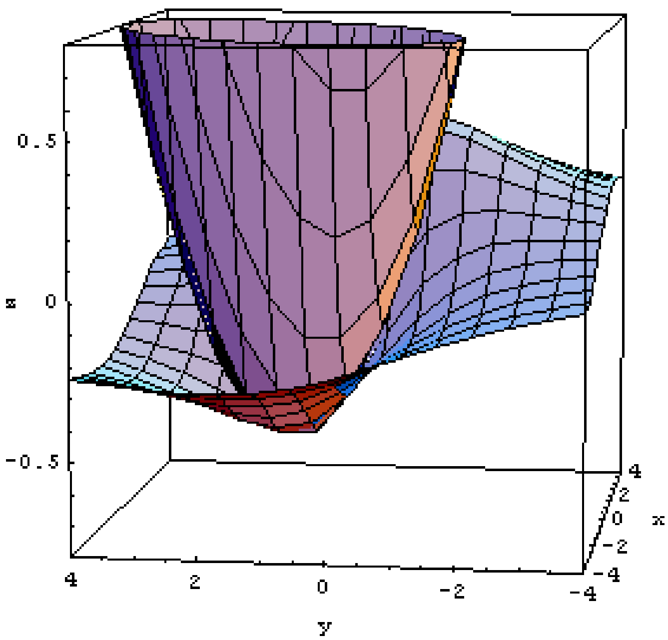


$$z = f[x,y] = \frac{x-y}{2+x^2+y^2}$$

$$z = Q[x,y] = \frac{-481 + 393\,x + 69\,x^2 - 763\,y - 78\,x\,y - 61\,y^2}{3375}$$

$Q[x,y]$ is centered at $\vec{P} = \{-3, -2\}$

When the point of expansion is closer to the minimum we can see that the minimum of the quadratic is close to the minimum of the surface. This is the idea used in Newton's method.



$$z = f[x,y] = \frac{x-y}{2+x^2+y^2}$$

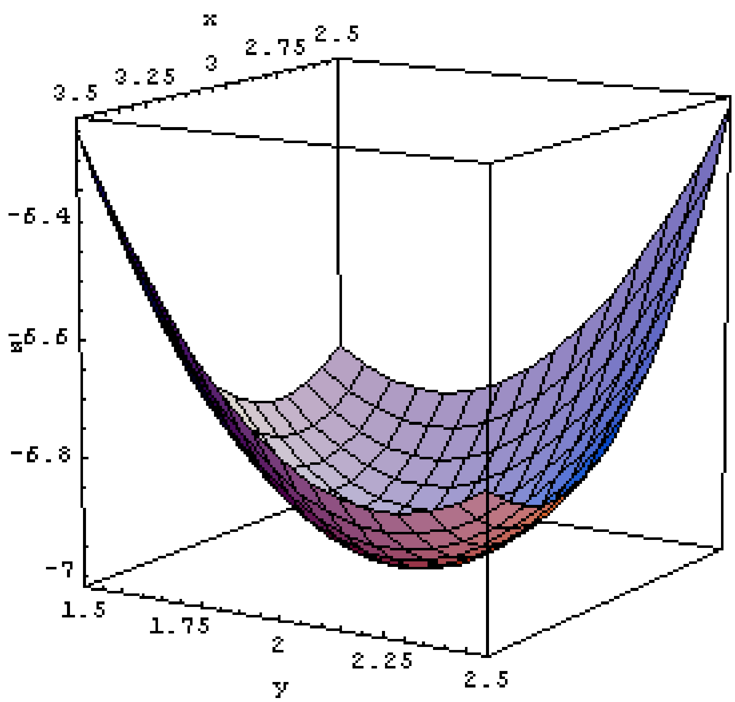
$$z = Q[x,y] = 0.0108293 + 0.576773\,x + 0.22383\,x^2 - 0.560112\,y - 0.195579\,x\,y + 0.190095\,y^2$$

$Q[x,y]$ is centered at $\vec{P} = \{-0.3, 0.2\}$

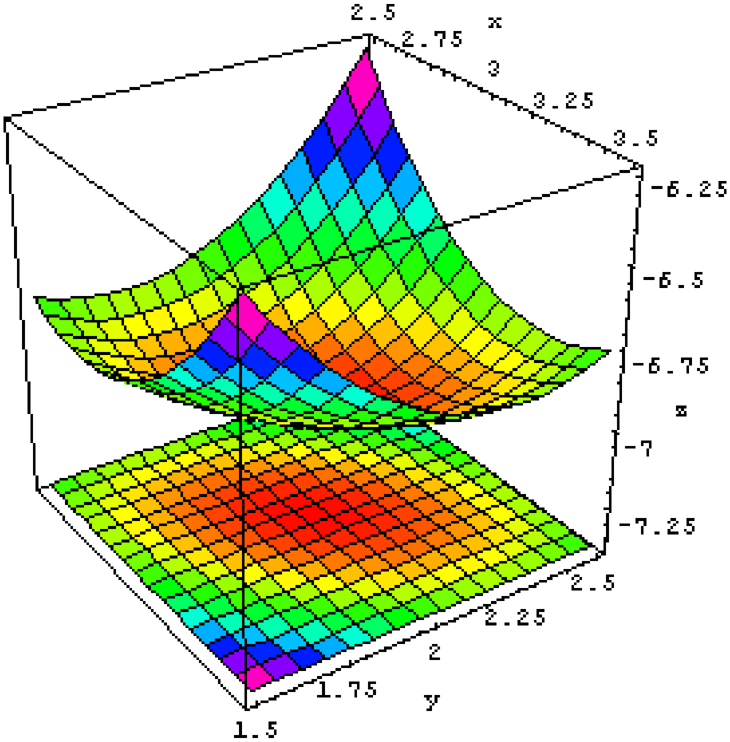
Example 3. Use the Newton search method to find the minimum of
$f(x, y) = x^2 - 4x + y^2 - y - x y$.

Solution 3.

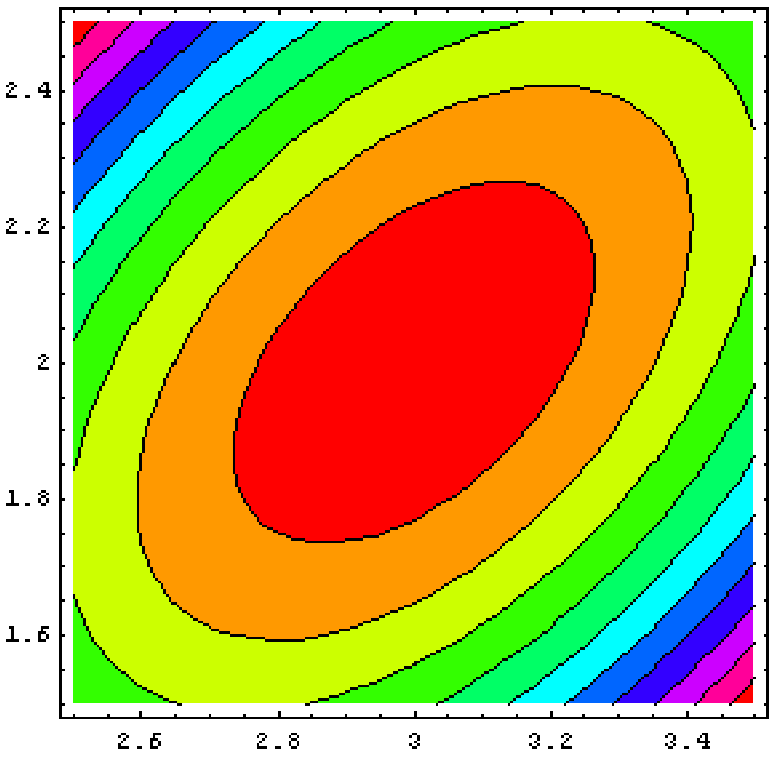
Enter the function $f(x, y) = x^2 - 4x + y^2 - y - x y$ and graph the surface $z = f(x, y)$.



$z = f[x,y] = -4x + x^2 - y - x y + y^2$



$z = f[x,y] = -4x + x^2 - y - xy + y^2$



$f[x,y] = -4x + x^2 - y - xy + y^2$

Execute the subroutine `NewtonSearch` and perform the iterations.

```
f[{ 0.000000000000,  0.000000000000}] =  0.000000000000
f[{ 3.000000000000,  2.000000000000}] = -7.000000000000
f[{ 3.000000000000,  2.000000000000}] = -7.000000000000
f[{ 3.000000000000,  2.000000000000}] = -7.000000000000
f[{ 3.000000000000,  2.000000000000}] = -7.000000000000
f[{ 3.000000000000,  2.000000000000}] = -7.000000000000
```

A local minimum of f[{x,y}] = $-4x + x^2 - y - xy + y^2$ is
f[{ 3.000000000000, 2.000000000000}] = -7.000000000000

Is the iterations surprising? It really converged in one iteration.
But this is to be expected because the approximation of a quadratic surface with a quadratic surface is exact.

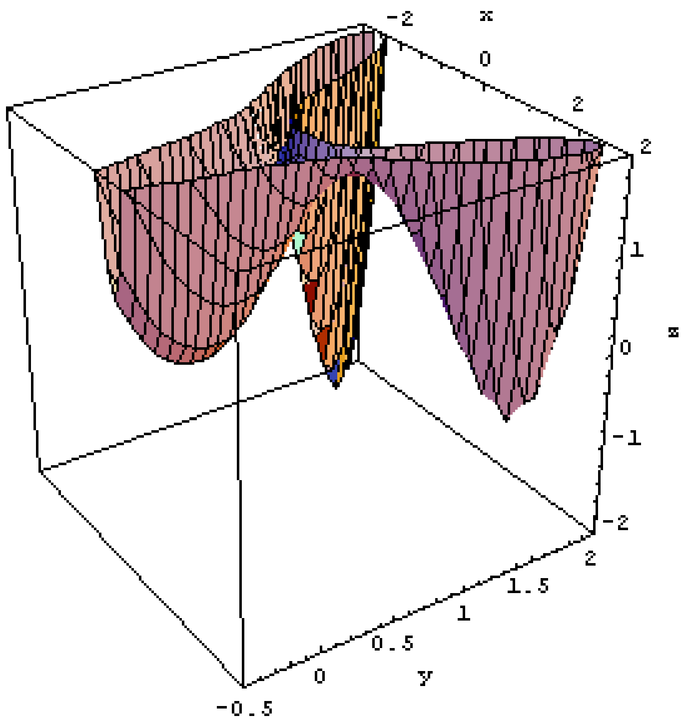
Let us compare this answer with *Mathematica*'s built in procedure **FindMinimum**.

A local minimum of f[{x,y}] = $-4x + x^2 - y - xy + y^2$ is
f[{ 3.000000000000, 2.000000000000}] = -7.000000000000

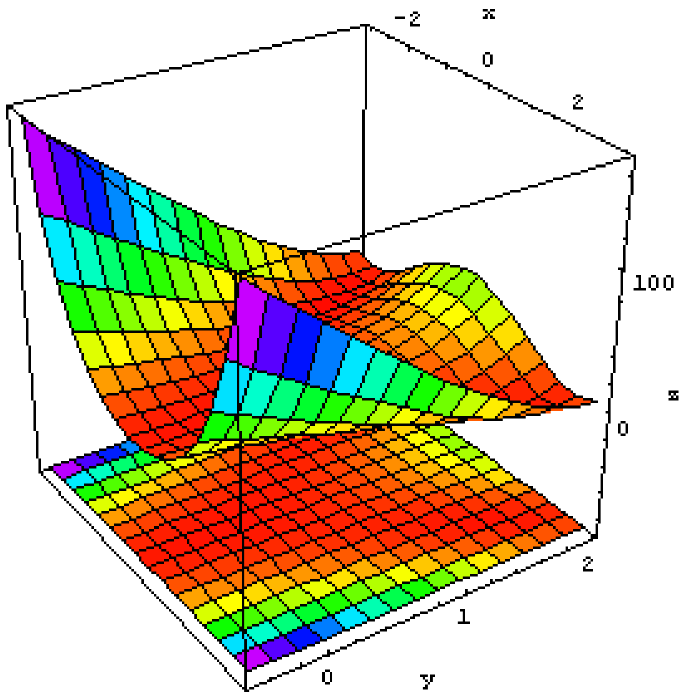
Example 4. Use the Newton method to find the minimum of $f(x, y) = 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4 + \frac{1}{4} x$. Looking at your graphs, estimate the location of the local minima.

Solution 4.

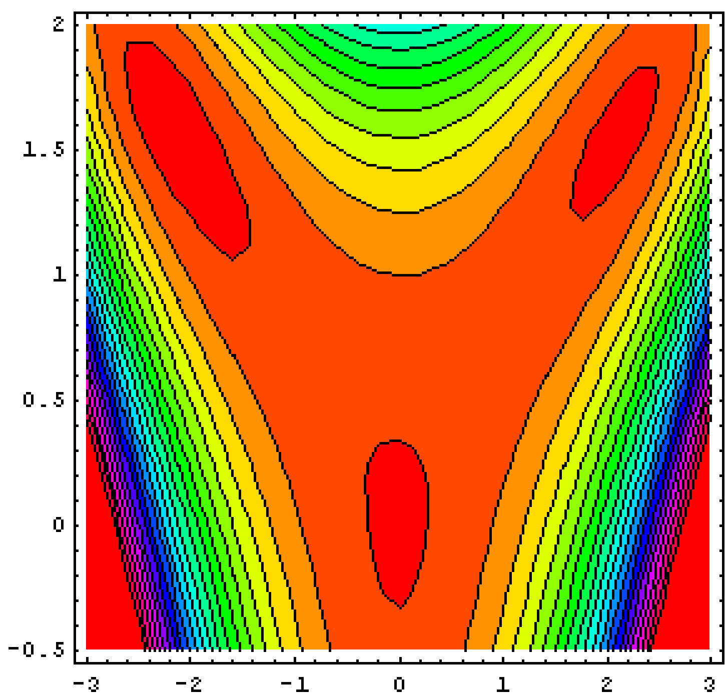
$$z = f[x,y] = \frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$$



$$z = f[x,y] = \frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$$



$$z = f[x,y] = \frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$$



$$f[x,y] = \frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$$

Looking at your graphs, estimate the location of the local minima.
Hint. The contour plot should be most useful.

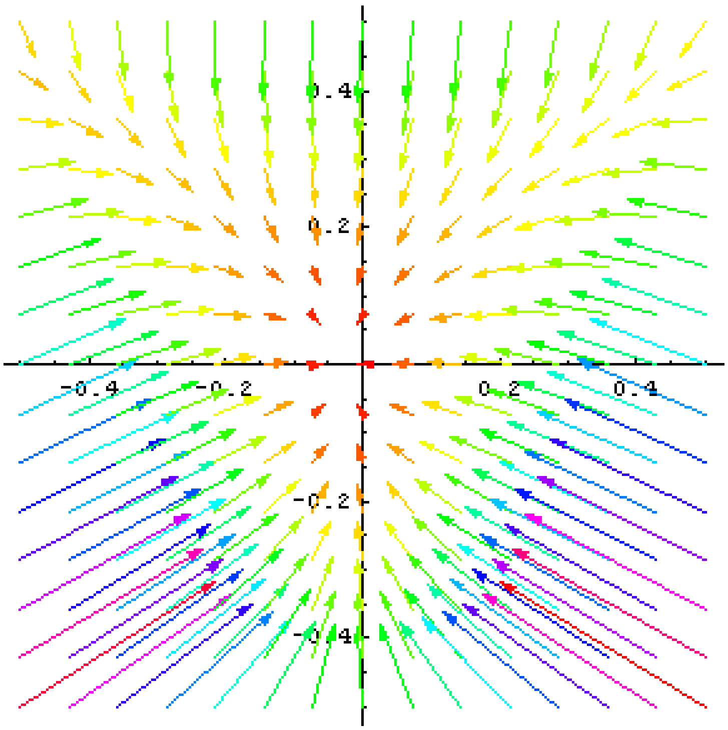
For the function $f(x, y) = 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4 + \frac{1}{4}x$, find the gradient vector and Hessian matrix.

$$f[\vec{X}] = \frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$$

$$\nabla f[\vec{X}] = \left\{\frac{1}{4}(1 + 40x + 16x^3 - 72xy), -9x^2 + 6y + 8y^3\right\}$$

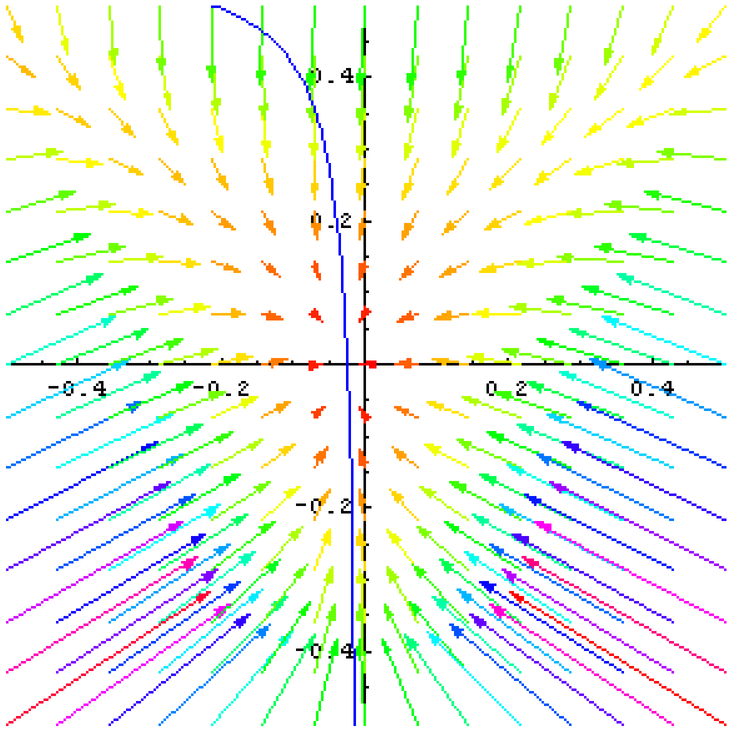
$$H[\vec{X}] = \begin{pmatrix} 2(5 + 6x^2 - 9y) & -18x \\ -18x & 6(1 + 4y^2) \end{pmatrix}$$

The minimum will occur where $-\nabla f(\vec{X})$ is the "zero vector." Let us investigate a neighborhood of the origin.



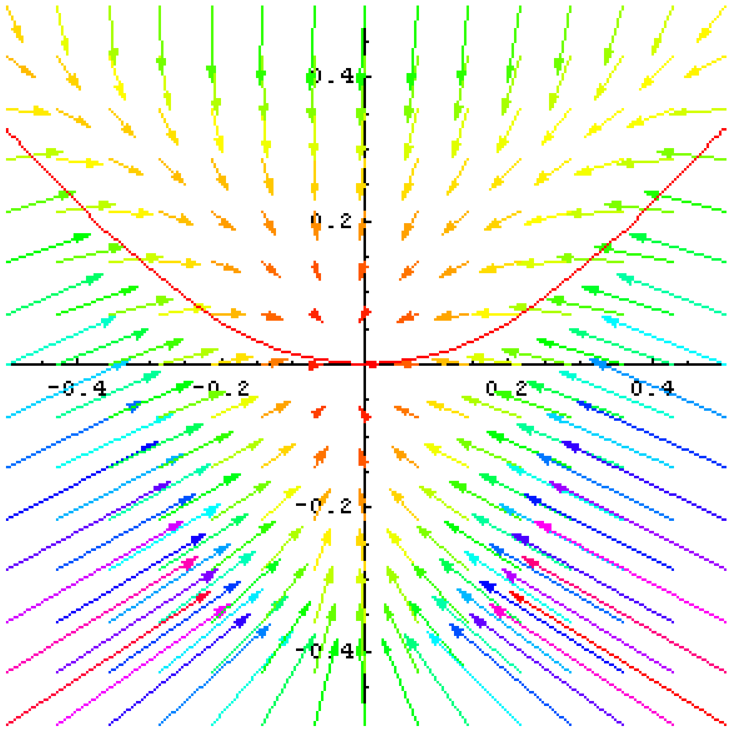
$$-\nabla f[\vec{X}] = \left\{\frac{1}{4}(-1 - 40x - 16x^3 + 72xy), 9x^2 - 6y - 8y^3\right\}$$

The following two graphs show the contour lines where the horizontal and vertical components of the gradient are zero.



The blue curve $\frac{1}{4}(-1 - 40x - 16x^3 + 72xy) == 0$

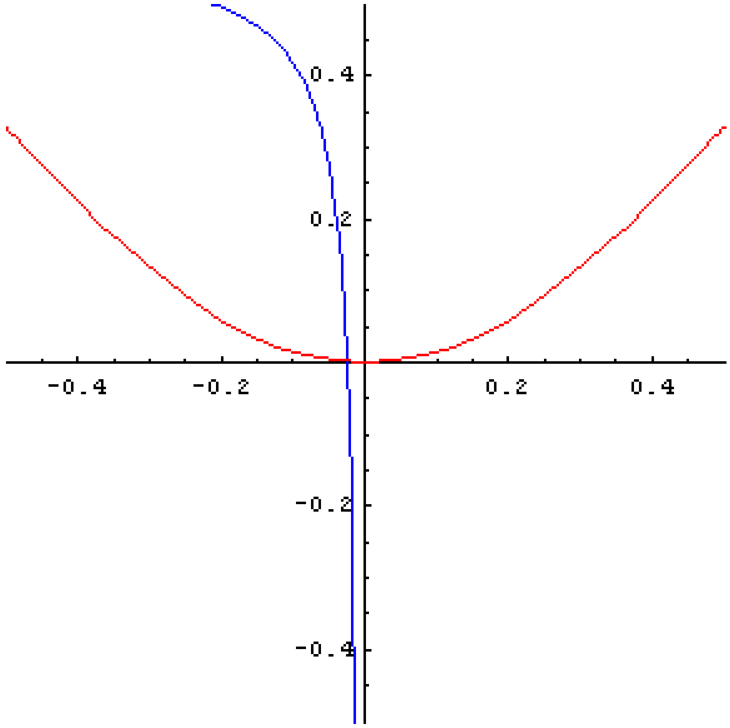
is where the horizontal component of $-\nabla f[\vec{X}]$ is zero.



The red curve $9x^2 - 6y - 8y^3 == 0$

is where the vertical component of $-\nabla f[\vec{X}]$ is zero.

Note. The red curve has a vertical asymptote x=0 that we want to remove.
The solution we seek is a point of intersection of the red curve and the blue curve.



The blue curve $\frac{1}{4}(1+40x+16x^3-72xy) == 0$

The red curve $-9x^2+6y+8y^3 == 0$

where the horizontal and vertical component of $-\nabla f[\vec{X}]$ are zero, repectively.

Case (i) Go for the minimum near {0.0, 0.0}

Execute the subroutine `NewtonSearch` and perform the iterations.

```
f[{ 0.000000000000,  0.000000000000}] =  0.000000000000
f[{-0.025000000000,  0.000000000000}] = -0.003124609375
f[{-0.025036032084,  0.000940202406}] = -0.003127252578
f[{-0.025036093327,  0.000940207845}] = -0.003127252578
f[{-0.025036093327,  0.000940207845}] = -0.003127252578
```

A local minimum of f[{x,y}] = $\frac{x}{4}+5x^2+x^4-9x^2y+3y^2+2y^4$ is

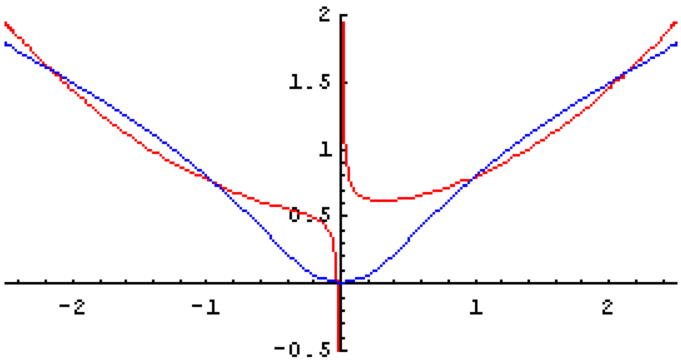
f[{-0.025036093327, 0.000940207845}] = -0.003127252578

Let us compare this answer with *Mathematica*'s built in procedure **FindMinimum**.

A local minimum of f[{x,y}] = $\frac{x}{4}+5x^2+x^4-9x^2y+3y^2+2y^4$ is

f[{-0.025036093029, 0.000940206139}] = -0.003127252578

Extraneous Solutions We could look at a larger plot where the horizontal and vertical components of the gradient are zero.

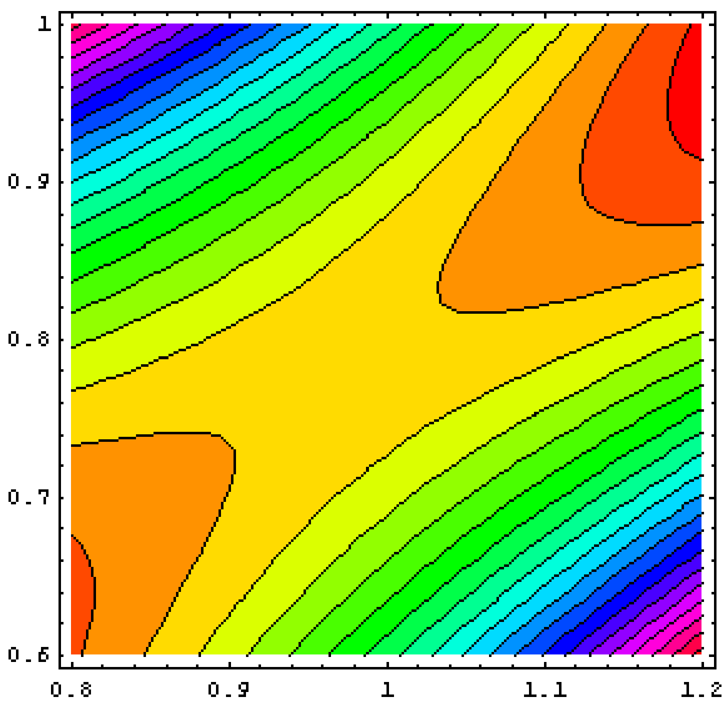


The blue curve $\frac{1}{4}(1 + 40\,x + 16\,x^3 - 72\,x\,y) == 0$

The red curve $-9\,x^2 + 6\,y + 8\,y^3 == 0$

where the horizontal and vertical component of $\nabla f[\vec{X}]$ are zero, repectively.

Note. The apparent solution near the point {1.0,0.8} turns out to be a saddle point and it could be located using Newton's method.
Unfortunately there is not way to determine this fact.



$f[x,y] = \frac{x}{4} + 5\,x^2 + x^4 - 9\,x^2\,y + 3\,y^2 + 2\,y^4$

We could perform the Newton iteration near this point. But keep in mind that a saddle point is located, not a minimum.

```
f[{ 1.000000000000,  0.800000000000}] =  1.789200000000
f[{ 0.968601543943,  0.778409540776}] =  1.792682331584
f[{ 0.968535517874,  0.778353659905}] =  1.792682348042
f[{ 0.968535517936,  0.778353659028}] =  1.792682348042
f[{ 0.968535517936,  0.778353659028}] =  1.792682348042
f[{ 0.968535517936,  0.778353659028}] =  1.792682348042
```

Further investigations are need to rule out this "extraneous solution."

```
f[{1., 0.8}] = 1.7892
But nearby points are not all larger.
f[{1.1, 0.9}] = 1.7303
f[{0.9, 0.9}] = 2.1123
f[{1.1, 0.7}] = 2.1163
f[{0.9, 0.7}] = 1.7783
 Therefore,
f[{1., 0.8}] = 1.7892  is NOT a local minimum.
```

Similarly, there is an "extraneous solution" near the point `{1.0,0.8}` which also turns out to be a saddle point.

Case (ii) Go for the minimum near `{2, 1.5}`

Execute the subroutine `NewtonSearch` and perform the iterations.

```
f[{ 2.000000000000,  1.500000000000}] = -0.625000000000
f[{ 2.186170212766,  1.611702127660}] = -0.752884718060
f[{ 2.149904635808,  1.588649103038}] = -0.763658971595
f[{ 2.148215779408,  1.587537848146}] = -0.763680059087
f[{ 2.148212130336,  1.587535403985}] = -0.763680059186
f[{ 2.148212130319,  1.587535403973}] = -0.763680059186
```

A local minimum of f[{x,y}] = $\frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$ is

```
f[{ 2.148212130319,  1.587535403973}] = -0.763680059186
```

Let us compare this answer with *Mathematica*'s built in procedure **FindMinimum**.

A local minimum of f[{x,y}] = $\frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$ is

```
f[{ 2.148211877409,  1.587535430391}] = -0.763680059184
```

Observation. Even *Mathematica* is having a hard time finding the minimum of

$f(x, y) = \frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$.

Since the function is "flat" near the minimum, the best way to achieve better accuracy is to increase the `WorkingPrecision`, i.e. use extended precision in the numerical computations.

A local minimum of f[{x,y}] = $\frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$ is

f[{ 2.148212130319, 1.587535403973}] = -0.763680059186

Case (iii) Go for the minimum near {-2, 1.5}

Execute the subroutine `NewtonSearch` and perform the iterations.

```
f[{-2.000000000000,  1.500000000000}] = -1.625000000000
f[{-2.239361702128,  1.643617021277}] = -1.818945395233
f[{-2.185339669518,  1.609338564133}] = -1.846180011106
f[{-2.181767365472,  1.606986963856}] = -1.846282957671
f[{-2.181751873418,  1.606976562851}] = -1.846282959604
f[{-2.181751873124,  1.606976562652}] = -1.846282959604
```

A local minimum of f[{x,y}] = $\frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$ is

f[{-2.181751873124, 1.606976562652}] = -1.846282959604

Let us compare this answer with *Mathematica*'s built in procedure **FindMinimum**.

A local minimum of f[{x,y}] = $\frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$ is

f[{-2.181751346205, 1.606976621830}] = -1.846282959597

Observation. Even *Mathematica* is having a hard time finding the minimum of

$f(x, y) = \frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$.

Since the function is "flat" near the minimum, the best way to achieve better accuracy is to increase the `WorkingPrecision`, i.e. use extended precision in the numerical computations.

A local minimum of f[{x,y}] = $\frac{x}{4} + 5x^2 + x^4 - 9x^2 y + 3y^2 + 2y^4$ is

f[{-2.181751873124, 1.606976562652}] = -1.846282959604