Finding a local minimum of a multivariate function

Our interest today will be finding a (local) minimum of a function $f\colon U \to \mathbb{R}$, where $U \subseteq \mathbb{R}^n$. (Of course, we already know how to do that with appropriate use of Newton's iteration.) Using the *gradient descent method* we find a minimum of a function $f\colon U \to \mathbb{R}$ by starting with some initial guess $\mathbf{x}^{(0)}$, and then continue iteratively:

$$\begin{aligned}
\mathbf{x}^{(1)} &= \mathbf{x}^{(0)} - h \operatorname{grad} f(\mathbf{x}^{(0)}),\\
\mathbf{x}^{(2)} &= \mathbf{x}^{(1)} - h \operatorname{grad} f(\mathbf{x}^{(1)}),\\
&\vdots\\
\mathbf{x}^{(k+1)} &= \mathbf{x}^{(k)} - h \operatorname{grad} f(\mathbf{x}^{(k)}).
\end{aligned}$$

Here, the step $h > 0$ is a carefully chosen (small) real number. (The convergence of such method depends on f, initial guess $\mathbf{x}^{(0)}$, and the choice of h.)

1. A function f is given by

$$f(x, y) = (1 - x)^2 + 4(y - x^2)^2.$$

   (a) Find the minimum of f.

   (b) Find (an approximation for) the minimum of the function f using the gradient descent method. Write an octave function `x = gradmet(gradf, h, x0, tol, maxit)`, which runs this method for the function f with gradient `gradf`, step size `h`, and initial guess `x0`. (We use `maxit` to limit the maximum allowed number of iterations, and `tol` to prescribe desired accuracy.)

2. Suppose we are given two circles, K and L, the first one with origin at (a, b) and radius r, the second one with origin at (a', b') and radius r'. We'd like to find the distance d between these two circles and points $P \in K$ and $Q \in L$ at this distance.

   (a) Express the distance d between these two circles analytically. (As a comparison with the method below.)

   (b) Write down the parametrizations $\mathbf{p}$ and $\mathbf{q}$ of circles K and L.

   (c) Let $f(t, u) = \|\mathbf{p}(t) - \mathbf{q}(u)\|^2$. Express $\operatorname{grad} f$ using parametrizations $\mathbf{p}$ and $\mathbf{q}$ (and defivatives $\dot{\mathbf{p}}$ and $\dot{\mathbf{q}}$).

   (d) Write an octave function `[d, T] = razdaljaK(K)`, which uses the gradient descent method to find the minimum of the function f, ie. the distance between these two circles. The input K is a 3×2 matrix with first column $[a, b, r]^\mathsf{T}$ and second column $[a', b', r']^\mathsf{T}$. The function should return the distance d and a 2×2 matrix T containing the spatial vectors of P and Q, ie. $T = [\mathbf{r}_P, \mathbf{r}_Q]$.

   (e) Can you use a similar method to find the points on K and L, which are farthest apart? Can you use a similar method to find the distance between two ellipses?

3. Can you find the distance of the previous exercise using the Newton's or Gauss–Newton iteration? Experiment and compare with gradient descent.

4. With some ingenuity we can use the gradient descent to solve systems of nonlinear equations. Instead of solving the system $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ we find the minimum of the function $f(\mathbf{x}) = \mathbf{F}(\mathbf{x})^\top \mathbf{F}(\mathbf{x}) = ||\mathbf{F}(\mathbf{x})||^2$.

   (a) Verify that
   $$\operatorname{grad} f(\mathbf{x}) = 2J\mathbf{F}(\mathbf{x})^\top \mathbf{F}(\mathbf{x}),$$
   holds. One step of gradient descent for the function $f = \mathbf{F}^\top \mathbf{F}$ is therefore
   $$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - 2hJ\mathbf{F}(\mathbf{x}^{(k)})^\top \mathbf{F}(\mathbf{x}^{(k)}).$$
   (Compare this with one step of Newton's iteration.)

   (b) Use the gradient descent to find at least one solution of the nonlinear system
   $$\begin{aligned} x_1^2 - x_2^2 - 1 &= 0, \\ x_1 + x_2 - x_1 x_2 - 1 &= 0. \end{aligned}$$