Let the objective function be given

$$f(x) = (x_1 - x_2)^2 + (x_1^2 - x_2 + 2)^2.$$

Find its gradient

$$f'(x) = \begin{pmatrix} 2(x_1 - x_2) + 4x_1(x_1^2 - x_2 + 2) \\ -2(x_1 - x_2) + 2(x_1^2 - x_2 + 2) \end{pmatrix}.$$

Consider the steepest descent method.

Gradient steepest descent method.

Step 0. Select an initial point x_0, ε_1, and ε_2, $k = 0$.

$$x_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad f(x_0) = 4,\ \varepsilon_1 = 0.02,\ \varepsilon_2 = 0.05.$$

Step 1. Calculate the gradient at this point

$$f'(x_0) = \begin{pmatrix} 0 \\ -4 \end{pmatrix}, \quad g_0 = -f'(x_0) = \begin{pmatrix} 0 \\ 4 \end{pmatrix}.$$

Step 2. Create a function

$$\varphi(\alpha) = f(x_0 + \alpha g_0) = 16\alpha^2 + (2 - 4\alpha)^2.$$

Minimizing this function by the parameter α we find the value of the step size

$$\varphi'(\alpha) = 64\alpha - 16.$$

Therefore, $\alpha_0 = \dfrac{1}{4}$.

Step 3. Put

$$x_1 = x_0 + \alpha_0 g_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

$$f(x_1) = 2. \quad f'(x_1) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

Step 4. Check the stopping criteria

$$||x_1 - x_0|| = 1 > \varepsilon_1, \quad ||f'(x_1)|| = 2 > \varepsilon_2.$$

The stopping criteria are not performed, so put $k = 1$, and go to Step 1.

The following iterations of the solution are summarized in a table.

Solving the problem by using the steepest descent method.

Table

k	x_k	$f(x_k)$	$f'(x_k)$	$\|\|x_{k+1} - x_k\|\|$	$\|\|f'(x_k)\|\|$
0	$(0, 0)^T$	4	$(0, -4)^T$	1	4
1	$(0, 1)^T$	28	$(-2, 0)^T$	0.313	2
2	$(0.313, 1.000)^T$	1.678	$(0.000, -0.822)^T$	0.2050	0.668
3	$(0.313, 1.205)^T$	1.593	$(-0.668, 0.000)^T$	0.0988	0.341
4	$(0.412, 1.205)^T$	1.559	$(-0.000, -0.341)^T$	0.0852	0.311
5	$(0.412, 1.291)^T$	1.545	$(-0.311, -0.000)^T$	0.0438	0.163
6	$(0.455, 1.291)^T$	1.538	$(-0.000, -0.163)^T$	0.0409	0.156
7	$(0.455, 1.331)^T$	1.535	$(-0.156, -0.000)^T$	0.0214	0.0829
8	$(0.477, 1.331)^T$	1.533	$(-0.0000, -0.0829)^T$	0.0207	0.0810
9	$(0.477, 1.352)^T$	1.5322	$(-0.0810, -0.000)^T$	0.0110	0.0431
10	$(0.488, 1.352)^T$	1.5317	$(-0.0000, -0.0431)^T$		

So now

$$||x_{10} - x_9|| \leq \varepsilon_1, \quad ||f'(x_9)|| \leq \varepsilon_2.$$

Let's consider some conjugate gradient methods.

The Fletcher-Reeves conjugate gradient method.

Step 0. Select an initial point x_0, ε_1, and ε_2.

$$x_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad f(x_0) = 4,\ \varepsilon_1 = 0.02,\ \varepsilon_2 = 0.05.$$

Step 1. Calculate the gradient at the initial point

$$f'(x_0) = \begin{pmatrix} 0 \\ -4 \end{pmatrix}, \quad g_0 = -f'(x) = \begin{pmatrix} 0 \\ 4 \end{pmatrix}.$$

Step 2. Create a function

$$\varphi(\alpha) = f(x_0 + \alpha g_0) = 16\alpha^2 + (2 - 4\alpha)^2.$$

Minimizing this function by the parameter α we find the value of the step size

$$\varphi'(\alpha) = 64\alpha - 16.$$

Therefore, $\alpha_0 = \dfrac{1}{4}$.

You can see that the first iteration of this method coincides with the first iteration of the steepest descent method, since the descent directions coincide.

The differences start from the second iteration.

Step 3. Set

$$x_1 = x_0 + \alpha_0 g_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$f(x_1) = 2,\ f'(x_1) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

Step 4. The stopping criteria are not performed

$$||x_1 - x_0|| > \varepsilon_1,\ ||f'(x_1)|| > \varepsilon_2.$$

Step 5. Calculate the new descent direction

$$\beta_0 = \frac{||f'(x_1)||^2}{||f'(x_0)||^2} = \frac{4}{16} = \frac{1}{4},$$

$$g_1 = -f'(x_1) + \beta_0 g_0 = -\begin{pmatrix} -2 \\ 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

.

Step 6. Put $k = 1$, go to Step 2.

Step 2. Create the function

$$\varphi(\alpha) = f(x_1 + \alpha g_1) = (\alpha - 1)^2 = (4\alpha^2 - \alpha + 1)^2.$$

Minimizing this function by the parameter α we find the value of the step size $\alpha_1 \approx 0.224$.

Step 3. Set

$$x_2 = x_1 + \alpha_1 g_1 \approx \begin{pmatrix} 0 \\ 1 \end{pmatrix} + 0.224 \begin{pmatrix} 2 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.449 \\ 1.224 \end{pmatrix},$$

$$f(x_2) \approx 1.556, \quad f'(x_2) \approx \begin{pmatrix} 0.201 \\ -0.402 \end{pmatrix}.$$

Step 4. The stopping criteria are not performed.

$$||x_2 - x_1|| \approx 0.501 > \varepsilon_1,\ ||f'(x_2)|| \approx 0.450 > \varepsilon_2.$$

Step 5. Calculate the new descent direction

$$\beta_1 = \frac{||f'(x_2)||^2}{||f'(x_1)||^2} \approx 0.0506,$$

$$g_2 = -f'(x_2) + \beta_1 g_1 = -\begin{pmatrix} 0.201 \\ -0.402 \end{pmatrix} + 0.0506 \begin{pmatrix} 2 \\ 1 \end{pmatrix} \approx \begin{pmatrix} -0.100 \\ 0.453 \end{pmatrix}.$$

.

Step 6. Set $k = 2$, go to Step 2.

The following iterations of the solution are summarized in a table.
Solving the problem using the Fletcher-Reeves method.

Table

k	x_k	$f(x_k)$	$f'(x_k)$	$\|\|x_{k+1} - x_k\|\|$	$\|\|f'(x_{k+1})\|\|$
0	$(0,0)^T$	4	$(0,-4)^T$	1	2
1	$(0,1)^T$	2	$(-2,0)^T$	0.501	0.450
2	$(0.449, 1.224)^T$	1.556	$(0.201, -0.402)^T$	0.0761	0.202
3	$(0.432, 1.299)^T$	1.540	$(-0.198, -0.044)^T$	0.0776	0.117
4	$(0.494, 1.346)^T$	1.532	$(-0.0710, 0.138)^T$	0.0211	0.0211
5	$(0.492, 1.367)^T$	1.5314	$(-0.0256, -2.144 \times 10^{-3})^T$	5.402×10^{-3}	0.0162
6	$(0.497, 1.368)^T$	1.5313	$(-5.362 \times 10^{-3}, -0.0153)^T$		

The Polak-Ribière conjugate gradient method.

Step 0. Select an initial point x_0, ε_1, and ε_2.

$$x_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad f(x_0) = 4, \ \varepsilon_1 = 0.02, \ \varepsilon_2 = 0.05.$$

Step 1. Calculate gradient at the initial point

$$f'(x_0) = \begin{pmatrix} 0 \\ -4 \end{pmatrix}, \quad g_0 = -f'(x) = \begin{pmatrix} 0 \\ 4 \end{pmatrix}.$$

Step 2. Create the function

$$\varphi(\alpha) = f(x_0 + \alpha g_0) = 16\alpha^2 + (2 - 4\alpha)^2.$$

Minimizing this function by the parameter α we find the value of the step size

$$\varphi'(\alpha) = 64\alpha - 16.$$

Therefore, $\alpha_0 = \dfrac{1}{4}$.

Step 3. Set

$$x_1 = x_0 + \alpha_0 g_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$f(x_1) = 2, \ f'(x_1) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

Step 4. The stopping criteria are not performed

$$||x_1 - x_0|| > \varepsilon_1, \ ||f'(x_1)|| > \varepsilon_2$$

Step 5. Calculate the new descent direction

$$f'(x_0) = \begin{pmatrix} 0 \\ -4 \end{pmatrix}, \quad f'(x_1) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}$$

$$f'(x_1) - f'(x_0) = \begin{pmatrix} -2 \\ 4 \end{pmatrix},$$

$$\beta_0 = \frac{\langle f'(x_1), (f'(x_1) - f'(x_0))\rangle}{||f'(x_0)||^2} = \frac{4}{16} = \frac{1}{4},$$

$$g_1 = -f'(x_1) + \beta_0 g_0 = -\begin{pmatrix} -2 \\ 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

.

Step 6. Set $k = 1$, go to Step 2.

Step 2. Create the function

$$\varphi(\alpha) = f(x_1 + \alpha g_1) = (\alpha - 1)^2 = (4\alpha^2 - \alpha + 1)^2.$$

Minimizing this function by the parameter α we find the value of the step size $\alpha_1 \approx 0.224$.

Step 3. Set

$$x_2 = x_1 + \alpha_1 g_1 \approx \begin{pmatrix} 0 \\ 1 \end{pmatrix} + 0.224 \begin{pmatrix} 2 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.449 \\ 1.224 \end{pmatrix},$$

$$f(x_2) \approx 1.556, \ f'(x_2) = \begin{pmatrix} 0.201 \\ -0.402 \end{pmatrix}.$$

Step 4. The stopping criteria are not performed

$$||x_2 - x_1|| \approx 0.501 > \varepsilon_1, \ ||f'(x_2)|| \approx 0.450 > \varepsilon_2.$$

Step 5. Calculate the new descent direction

$$f'(x_1) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \quad f'(x_2) \approx \begin{pmatrix} 0.201 \\ -0.402 \end{pmatrix}$$

$$f'(x_2) - f'(x_1) \approx \begin{pmatrix} 2.201 \\ -0.402 \end{pmatrix},$$

$$\beta_1 = \frac{\langle f'(x_2), (f'(x_2) - f'(x_1)) \rangle}{||f'(x_1)||^2} \approx \frac{0.604}{4} = 0.151,$$

$$g_2 = -f'(x_2) + \beta_1 g_1 \approx -\begin{pmatrix} 0.201 \\ -0.402 \end{pmatrix} + 0.151 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.101 \\ 0.554 \end{pmatrix}.$$

Starting from this point the solutions of the problem by the methods of Fletcher- Reeves and Polak-Ribière begin to differ.

Step 6. Set $k = 2$, go to Step 2.

The following iterations of the solution are summarized in a table.

Solving the problem by using the Polak-Ribière method.

Table

k	x_k	$f(x_k)$	$f'(x_k)$	$\|\|x_{k+1} - x_k\|\|$	$\|\|f'(x_{k+1})\|\|$
0	$(0,0)^T$	4	$(0,-4)^T$	1	2
1	$(0,1)^T$	2	$(-2,0)^T$	0.501	0.450
2	$(0.449, 1.224)^T$	1.556	$(0.201, -0.402)^T$	0.131	0.121
3	$(0.472, 1.353)^T$	1.533	$(-0.119, 0.0218)^T$	0.0347	$4,756 \times 10^{-3}$
4	$(0.500, 1.374)^T$	1.531	$(2.866 \times 10^{-3}, 7.080 \times 10^{-3})^T$	1.230×10^{-3}	5.783×10^{-3}
5	$(0.500, 1.375)^T$	1.531	$(-5.659 \times 10^{-3}, 1.188 \times 10^{-3})^T$		

Consider the Quasi-Newton method

The Davidon-Fletcher-Powell method

Step 0. Select an initial point x_0, ε_1, and ε_2.

Step 1. Put k=0, $H_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Calculate $f(x_0)$=4.

Step 2. Find the gradient and the descent direction

$$f'(x_0) = \begin{pmatrix} 0 \\ -4 \end{pmatrix}, \quad p_0 = -H_0 f'(x_0) = \begin{pmatrix} 0 \\ 4 \end{pmatrix}.$$

Step 3. Create the function

$$\varphi(\alpha) = f(x_0 + \alpha p_0) = (\alpha - 1)^2 = (4\alpha^2 - \alpha + 1)^2.$$

Minimizing this function by the parameter α we find the value of the step size $\alpha_0 = \frac{1}{4}$.

Step 4. Set

$$x_1 = x_0 + \alpha_0 p_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 0 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$f(x_1) = 2, \quad f'(x_1) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}.$$

Step 5. The stopping criteria are not performed

$$||x_1 - x_0|| = 1 > \varepsilon_1, \; ||f'(x_1)|| = 2 > \varepsilon_2.$$

Step 6. Calculate new matrix H_k

$$\Delta_0 = x_1 - x_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$y_0 = f'(x_1) - f'(x_0) = \begin{pmatrix} -2 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ -4 \end{pmatrix} = \begin{pmatrix} -2 \\ 4 \end{pmatrix},$$

$$H_{01} = \frac{\Delta_0 (\Delta_0)^T}{(\Delta_0)^T \Delta_0} - \frac{H_0 y_0 (y_0)^T (H_0)^T}{(y_0)^T (H_0)^T y_0} =$$

$$= \frac{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \end{pmatrix}}{\begin{pmatrix} 0 & 1 \end{pmatrix}\begin{pmatrix} -2 \\ 4 \end{pmatrix}} - \frac{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -2 \\ 4 \end{pmatrix}\begin{pmatrix} -2 & 4 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}{\begin{pmatrix} -2 & 4 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} -2 \\ 4 \end{pmatrix}} = \begin{pmatrix} -0.2 & 0.4 \\ 0.4 & -0.55 \end{pmatrix}.$$

$$H_1 = H_0 + H_{01} = \begin{pmatrix} 0.8 & 0.4 \\ 0.4 & 0.45 \end{pmatrix}$$

Step 7. Set k=1, goto Step 2.

Step 2. Find new descent direction

$$p_1 = -H_1 f'(x_1) = \begin{pmatrix} 1.6 \\ 0.8 \end{pmatrix},$$

Step 3. Create the function

$$\varphi(\alpha) = f(x_1 + \alpha p_0) = (0.8\alpha - 1)^2 + (2.56\alpha^2 - 0.8\alpha + 1)^2.$$

Minimizing this function by the parameter α we find the value of the step size $\alpha_1 \approx 0.280$.

Step 4. Set

$$x_2 = x_1 + \alpha_1 p_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + 0.280\begin{pmatrix} 1.6 \\ 0.8 \end{pmatrix} = \begin{pmatrix} 0,449 \\ 1.224 \end{pmatrix},$$

$$f(x_2) = 1.556, \quad f'(x_2) = \begin{pmatrix} 0.201 \\ -0.402 \end{pmatrix}.$$

Step 5. The stopping criteria are not performed

$$||x_1 - x_0|| = 0.501 > \varepsilon_1, \ ||f'(x_1)|| = 0.405 > \varepsilon_2.$$

Step 6. Calculate new matrix H_2

$$\Delta_1 = x_1 - x_0 = \begin{pmatrix} 0.449 \\ 0.224 \end{pmatrix},$$

$$y_1 = f'(x_1) - f'(x_0) = \begin{pmatrix} 0.201 \\ -0402 \end{pmatrix} - \begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 2.201 \\ -0.402 \end{pmatrix},$$

$$H_{11} = \frac{\Delta_1(\Delta_1)^T}{(\Delta_1)^T\Delta_1} - \frac{H_1 y_1 (y_1)^T (H_1)^T}{(y_1 * 0)^T (H_1)^T y_1} = \begin{pmatrix} -0.566 & -0.233 \\ -0.233 & -0.095 \end{pmatrix}.$$

$$H_2 = H_1 + H_{11} = \begin{pmatrix} 0.234 & 0.167 \\ 0.167 & 0.355 \end{pmatrix}$$

Step 7. Set k=2, goto Step 2.

Table

k	x_k	$f(x_k)$	$f'(x_k)$	H_k	$\|\|x_{k+1} - x_k\|\|$	$\|\|f'(x_{k+1})\|\|$
0	$(0,0)^T$	4	$(0,-4)^T$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	1	2
1	$(0,1)^T$	2	$(-2,0)^T$	$\begin{pmatrix} 0.8 & 0.4 \\ 0.4 & 0.45 \end{pmatrix}$	0.501	0.450
2	$(0.449, 1.224)^T$	1.556	$(0.201, -0.402)^T$	$\begin{pmatrix} 0.234 & 0.167 \\ 0.167 & 0.355 \end{pmatrix}$	0.131	0.121
3	$(0.472, 1.353)^T$	1.533	$(-0.119, 0.0218)^T$	$\begin{pmatrix} 0.246 & 0.241 \\ 0.241 & 485 \end{pmatrix}$	0.0347	$4,756 \times 10^{-3}$
4	$(0.500, 1.374)^T$	1.531	$(2.866 \times 10^{-3}, -3.79510^{-3})^T$	$\begin{pmatrix} 0.286 & 0.281 \\ 0.281 & 0.522 \end{pmatrix}$	1.230×10^{-3}	5.784×10^{-3}
5	$(0.500, 1.375)^T$	1.531	$(-5.661 \times 10^{-5}, 1.190 \times 10^{-5})^T$			