Gradient descent

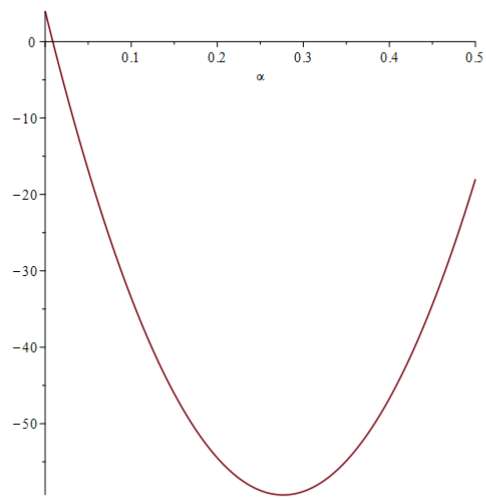
1. Apply gradient descent for finding the extrema of the function starting with the point $x = 0$ and $y = 0$:

$$f(x, y) = 3x^2 + 5y^2 + 4xy + 17x - 13y + 4$$

Answer: Let $\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}$ and let $\mathbf{u}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. The gradient of this function is:

$$\vec{\nabla} f(\mathbf{u}) = \begin{pmatrix} 6x + 4y + 17 \\ 4x + 10y - 13 \end{pmatrix}.$$

The negation of the gradient at the initial point is therefore $-\vec{\nabla} f(\mathbf{u}_0) = \begin{pmatrix} -17 \\ 13 \end{pmatrix}$. Thus, if we plot $f\left(\mathbf{u}_0 - \alpha_0 \vec{\nabla} f(\mathbf{u}_0)\right)$, we have



Now, if you were to evaluate f at this linear combination of vectors, you would find that this defines a function

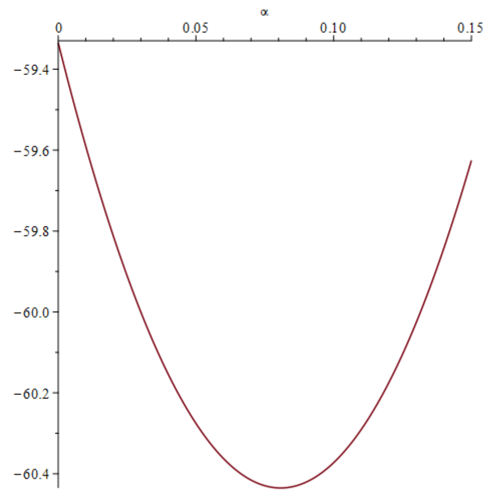
$$828\alpha_0^2 - 458\alpha_0 + 4,$$

which has a minimum at $\alpha_0 = \frac{229}{828} \approx 0.2765700483091787$. Thus, our next approximation is

$$\mathbf{u}_1 \leftarrow \mathbf{u}_0 - \alpha_0 \vec{\nabla} f(\mathbf{u}_0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 0.2765700483091787 \begin{pmatrix} 17 \\ -13 \end{pmatrix} = \begin{pmatrix} -4.701690821256039 \\ 3.595410628019324 \end{pmatrix}$$

and $f(\mathbf{u}_1) = -59.33454106280195$.

The negation of the gradient at the next point is therefore $-\vec{\nabla} f(\mathbf{u}_1) = \begin{pmatrix} -3.17149758454107 \\ -4.14734299516908 \end{pmatrix}$. Thus, if we plot $f\left(\mathbf{u}_1 - \alpha_1 \vec{\nabla} f(\mathbf{u}_1)\right)$, we have



Now, if you were to evaluate f at this linear combination of vectors, you would find that this defines a function

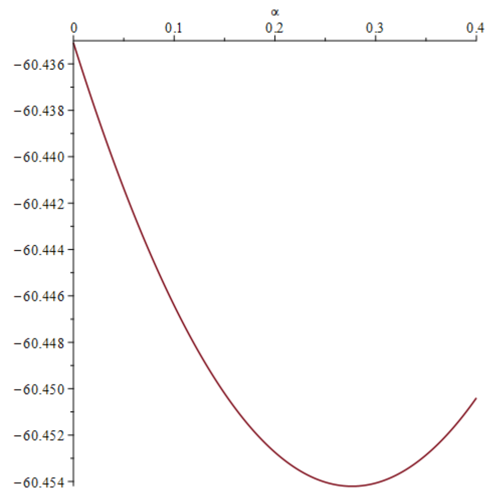
$$168.7906135499080\alpha_1^2 - 27.258850848327\alpha_1 - 59.33454106280198,$$

which has a minimum at $\alpha_1 \approx 0.080747531734837786$. Thus, our next approximation is

$$\begin{aligned} \mathbf{u}_2 \leftarrow \mathbf{u}_1 - \alpha_1 \vec{\nabla} f(\mathbf{u}_1) &= \begin{pmatrix} -4.701690821256039 \\ 3.595410628019324 \end{pmatrix} - 0.080747531734837786 \begin{pmatrix} 3.17149758454107 \\ 4.14734299516908 \end{pmatrix} \\ &= \begin{pmatrix} -4.957781423110731 \\ 3.260522917901651 \end{pmatrix} \end{aligned}$$

and $f(\mathbf{u}_2) = -60.43508352476719$.

Finally, the negation of the gradient at the second point is therefore $-\vec{\nabla} f\left(\mathbf{u}_2\right)=\begin{pmatrix} -0.29540313294221 \\ 0.22589651342641 \end{pmatrix}$. Thus, if we plot $f\left(\mathbf{u}_2-\alpha_2\vec{\nabla} f\left(\mathbf{u}_2\right)\right)$, we have



Now, if you were to evaluate f at this linear combination of vectors, you would find that this defines a function

$$0.2500130555997261\alpha_2^2 - 0.138292245730284\alpha_2 - 60.43508352476720$$

\which has a minimum at $\alpha_2 \approx 0.2765700483091802$. Thus, our third approximation is

$$\mathbf{u}_3 \leftarrow \mathbf{u}_2 - \alpha_2\vec{\nabla} f\left(\mathbf{u}_2\right) = \begin{pmatrix} -4.957781423110731 \\ 3.260522917901651 \end{pmatrix} - 0.2765700483091802\begin{pmatrix} 0.29540313294221 \\ -0.22589651342641 \end{pmatrix}$$
$$= \begin{pmatrix} -5.039481081859241 \\ 3.322999127532869 \end{pmatrix}$$

and $f(\mathbf{u}_3) = -60.45420727130844$.

2. Are consecutive gradients perpendicular to each other?

Answer: It appears yes, for if you take the inner product between them, that value is very close to zero.

3. You may have noticed that $\alpha_0 = \alpha_2$. Is this to be expected; that is, do you expect $\alpha_k = \alpha_{k+2}$ in general?

Answer: No, this is happening here simply because the function being minimize is a quadratic. As there are more direct methods for finding the minimum of a quadratic, this isn't even really that useful an observation.

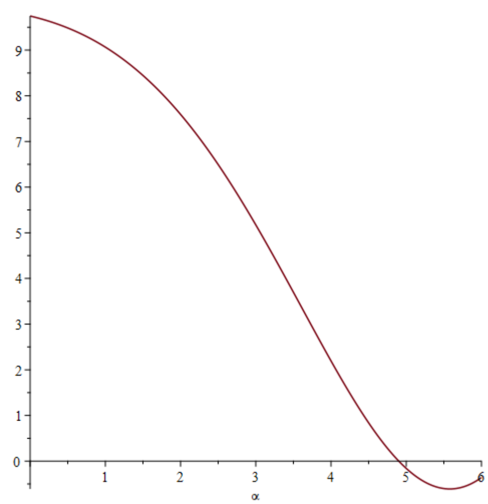
4. Apply gradient descent of the function starting with the point $x = 1$, $y = 1$ and $z = 1$:

$$f(x, y, z) = 4\cos(0.3xy) + 3\cos(0.2yz) + 3\cos(0.1xz)$$

Answer: Let $\mathbf{u} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ and let $\mathbf{u}_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. The gradient of this function is

$$\vec{\nabla} f(\mathbf{u}) = \begin{pmatrix} -1.2y\sin(0.3xy) - 0.3z\sin(0.1xz) \\ -0.6z\sin(0.2xz) - 1.2x\sin(0.3xy) \\ -0.3x\sin(0.1xz) - 0.6y\sin(0.2xz) \end{pmatrix}.$$

The negation of the gradient at the initial point is therefore $-\vec{\nabla} f(\mathbf{u}_0) = \begin{pmatrix} 0.3845742729876559 \\ 0.4738258464706442 \\ 0.1491516234710851 \end{pmatrix}$. Thus, if we plot $f\left(\mathbf{u}_0 - \alpha_0 \vec{\nabla} f(\mathbf{u}_0)\right)$, we have



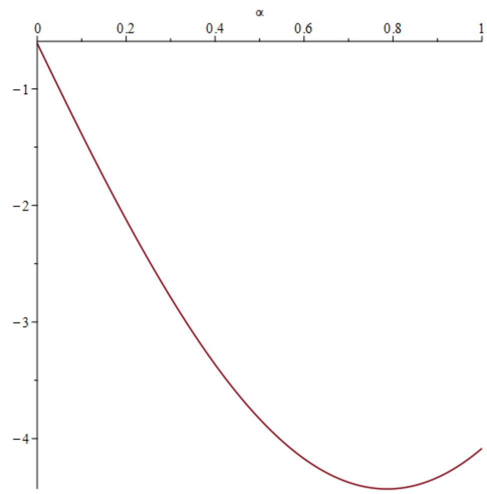
Using one-dimensional numerical optimization methods, we may determine that this has a minimum at $\alpha_0 \approx 5.579412899276763$. Thus, our next approximation is

$$\mathbf{u}_1 \leftarrow \mathbf{u}_0 - \alpha_0 \vec{\nabla} f(\mathbf{u}_0) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - 5.579412899276763 \begin{pmatrix} -0.3845742729876559 \\ -0.4738258464706442 \\ -0.1491516234710851 \end{pmatrix}$$

$$= \begin{pmatrix} 3.145698659437311 \\ 3.643670039809043 \\ 1.832178491942643 \end{pmatrix}$$

and $f(\mathbf{u}_1) = -0.609176731786268$.

The negation of the gradient at the next point is therefore $-\vec{\nabla} f\left(u_0\right) = \begin{pmatrix} -0.9799444827745933 \\ -0.035690900336148 \\ 2.640083285253116 \end{pmatrix}$. Thus, if we plot $f\left(\mathbf{u}_1 - \alpha_1 \vec{\nabla} f\left(\mathbf{u}_1\right)\right)$, we have



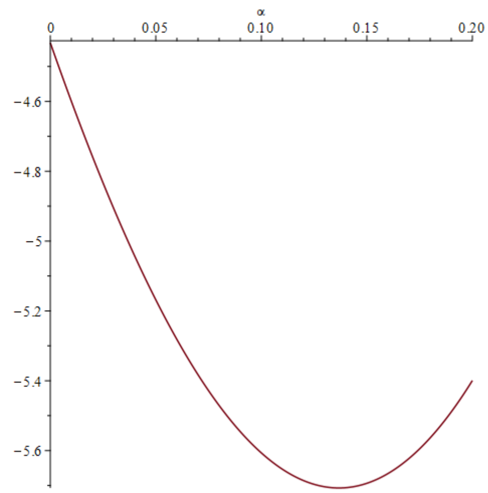
Using one-dimensional numerical optimization methods, we may determine that this has a minimum at $\alpha_1 \approx 0.7866853844524082$. Thus, our next approximation is

$$u_2 \leftarrow u_1 - \alpha_1 \vec{\nabla} f\left(u_1\right) = \begin{pmatrix} 3.145698659437311 \\ 3.643670039809043 \\ 1.832178491942643 \end{pmatrix} - 0.7866853844524082 \begin{pmatrix} 0.9799444827745933 \\ 0.035690900336148 \\ -2.640083285253116 \end{pmatrix}$$

$$= \begin{pmatrix} 2.374790657263764 \\ 3.615592530156648 \\ 3.909093426188367 \end{pmatrix}$$

and $f(\mathbf{u}_2) = -4.431836171000048$.

Finally, the negation of the gradient at the second point is therefore $-\vec{\nabla} f\left(\mathbf{u}_2\right)=\begin{pmatrix}-0.29540313294221\\ 0.22589651342641\end{pmatrix}$. Thus, if we plot $f\left(\mathbf{u}_2-\alpha_2\vec{\nabla} f\left(\mathbf{u}_2\right)\right)$, we have



Using one-dimensional numerical optimization methods, we may determine that this has a minimum at $\alpha_2 \approx 0.1367658073293575$. Thus, our third approximation is

$$u_3 \leftarrow u_2-\alpha_2\vec{\nabla} f\left(u_2\right)=\begin{pmatrix}2.374790657263764\\ 3.615592530156648\\ 3.909093426188367\end{pmatrix}-0.1367658073293575\begin{pmatrix}-3.264524355445020\\ -2.253846402564102\\ -1.242193553759089\end{pmatrix}$$

$$=\begin{pmatrix}2.821265966282553\\ 3.923841652999696\\ 4.078983030427552\end{pmatrix}$$

and $f(\mathbf{u}_3) = -5.707153277803242$.

5. This method appears to be converging much more slowly. Any ideas why?

Answer: This is a very highly non-linear function with many local peaks and troughs. Consequently, there is no real direct route from what is close to a local maximum (recall that $f(\mathbf{u}_0) = 9.746558185860226$, to a minimum. If, however, we were to start at a reasonable approximation of a minimum, we may be able to converge to a better approximation.

6. Instead of starting with the vector we did in Question 4, start with the vector $\mathbf{u}_0 = \begin{pmatrix} 4.75 \\ 2.25 \\ 6.75 \end{pmatrix}$. What is the sequence of approximations?

Answer: If we let $\mathbf{u}_0 = \begin{pmatrix} 4.75 \\ 2.25 \\ 6.75 \end{pmatrix}$ where $f(\mathbf{u}_0) = -9.969134842562037$, our sequence of iterations now proceed as follows:

1. $\mathbf{u}_1 = \begin{pmatrix} 4.597650581296476 \\ 2.276210979603908 \\ 6.774052543229699 \end{pmatrix}$ where $f(\mathbf{u}_1) = -9.993886012682701$
2. $\mathbf{u}_2 = \begin{pmatrix} 4.602120207739099 \\ 2.294328198638612 \\ 6.782620241447279 \end{pmatrix}$ where $f(\mathbf{u}_2) = -9.996748987558708$
3. $\mathbf{u}_3 = \begin{pmatrix} 4.586999613729395 \\ 2.282185635918663 \\ 6.816185017884426 \end{pmatrix}$ where $f(\mathbf{u}_3) = -9.998270666217952$