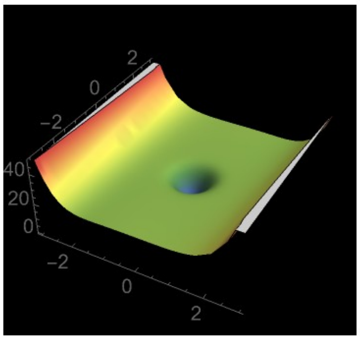
LCPB 20-21 exercise 1

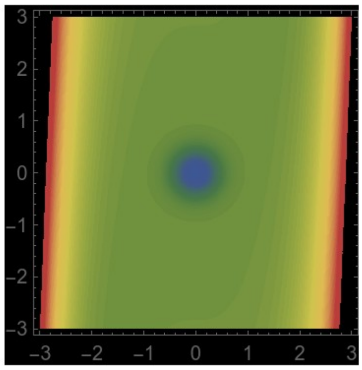
Consider notebook NB2 by Mehta et al., which can be found at this website:
http://physics.bu.edu/~pankajm/MLnotebooks.html

1. ~~Add the ADAmax algorithm (find its definition outside the review by Mehta)~~
2. Show a quantitative statistical comparison of the performance of different algorithms:
   - Vanilla gradient descent
   - Gradient descent with momentum
   - Nesterov (NAG)
   - RMSprop
   - ADAM
   - ~~ADAmax~~

   for this function on the right or for the function in the next page.

   Define a grid Q of initial points equally spaced in the square S=[-3,3]x[-3,3]. Perform a minimization starting from each of the points in Q, and compute the average value of the function vs time during these minimizations, for each method (with a good value of its own learning rate, chosen after some test). ~~Eventually (a) plot also the standard deviation around the average value; (b) plot data vs real CPU time rather than "*t*" of the iteration (it could be a better comparison because some methods are more complicated and use more CPU).~~





function

$$b\left(1-e^{-\frac{1}{2}w\left(x^2+y^2\right)}\right)+\frac{1}{2}q\left(-x^3+y\right)^2$$

gradient of the function, component x

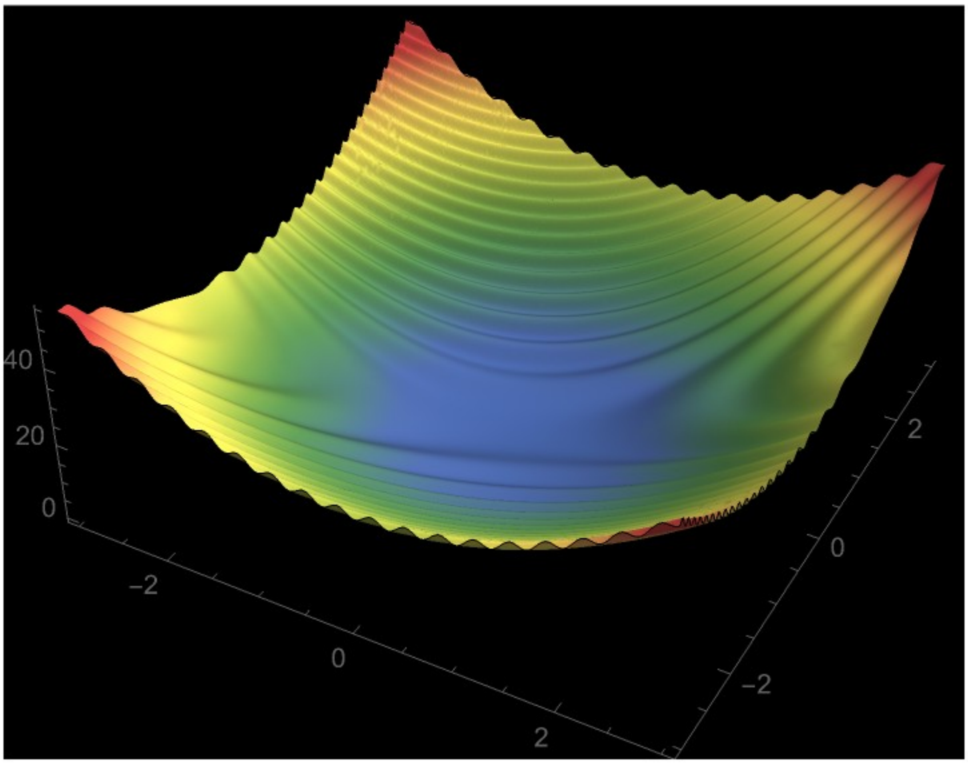
$$b\,e^{-\frac{1}{2}w\left(x^2+y^2\right)}\,w\,x-3\,q\,x^2\left(-x^3+y\right)$$

gradient of the function, component y

$$b\,e^{-\frac{1}{2}w\left(x^2+y^2\right)}\,w\,y+q\left(-x^3+y\right)$$

parameters: $w=10,\ q=\frac{1}{10},\ b=20$

3. OPTIONAL: For a simple function, show an example where ADAM algorithm starts to become unstable with respect to a minimum that was reached at some earlier iteration *t*. ~~Compare it with ADAmax behavior.~~



function

$$1 + \frac{1}{2} q \left(x^2 + y^2\right) - \mathrm{Cos}\left[2 \pi \left(x y - y^2\right)\right]$$

gradient of the function, component x

$$q x + 2 \pi y \,\mathrm{Sin}\left[2 \pi \left(x y - y^2\right)\right]$$

gradient of the function, component y

$$q y + 2 \pi (x - 2 y) \,\mathrm{Sin}\left[2 \pi \left(x y - y^2\right)\right]$$

parameters: $q=6$