

# 6.867 : Homework 1

Anonymous authors

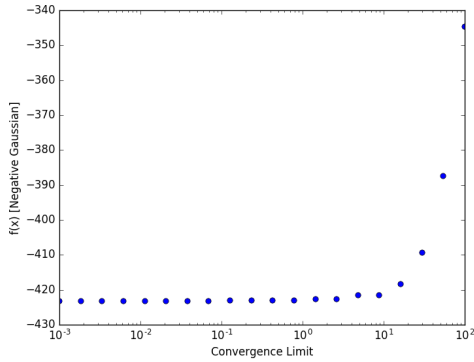


Fig. 2: The final value returned by GD can differ substantially if convergence limit is set too loosely.

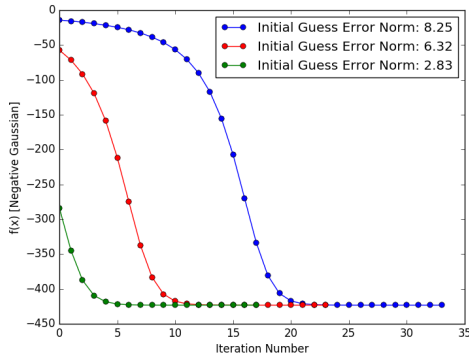


Fig. 3: When initial guess is far from true minimum, it takes many iterations (blue) to converge. As initial guess approaches final solution, number of iterations decreases.

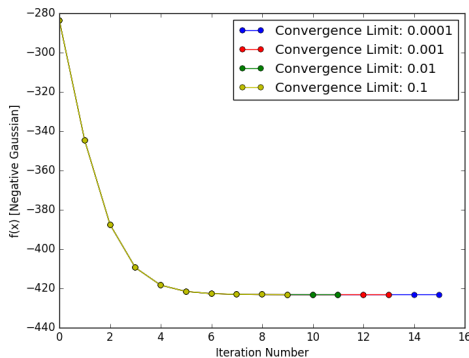


Fig. 1: As convergence limit,  $\epsilon$ , increases, the iteration stops earlier. From this view, the four solutions differ very little.

## I. GRADIENT DESCENT

In this section, we implemented Gradient Descent in Python to find the minimum of a function.

### A. Part 1

Our Gradient Descent (GD) implementation begins with an initial guess,  $x_0$ , of the argument  $x_{min}$  that minimizes the function,  $f(x)$ . The implementation is applied to two functions, a negative Gaussian and a quadratic bowl. The gradient,  $dfdx$ , can be calculated in closed form (for simple functions), and the next estimate of  $x_{min}$  is updated according to the rule  $x_{i+1} = x_i - \alpha \cdot \frac{\partial f(x_i)}{\partial x}$ . Here,  $\alpha$  is the step size, or learning rate, which controls how much the gradient affects the next estimate. The rule is applied until either a pre-determined maximum number of iterations,  $n_{iters}$ , occurs, or the function's value at the current iteration differs from

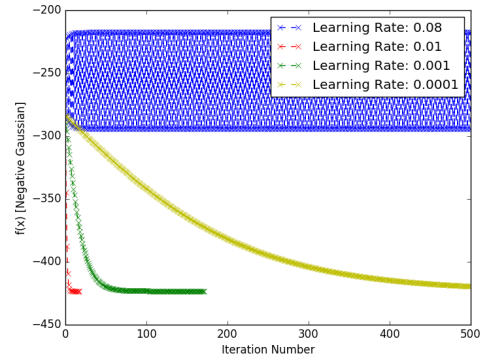


Fig. 4: If learning rate,  $\alpha$ , is too large (blue), the solution will oscillate or diverge each iteration. If  $\alpha$  is too small, it will take many iterations to converge (yellow). Convergence takes only a few iterations for proper  $\alpha$  (red).

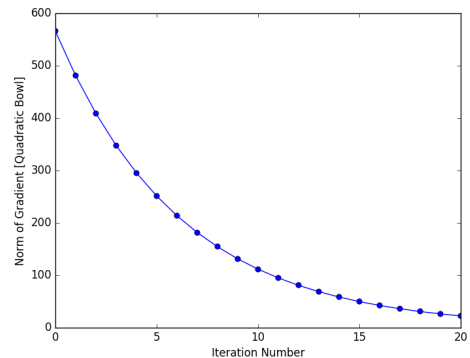


Fig. 5: Norm of gradient decreases with each iteration in GD. This example minimizes the Quadratic Bowl function.

the previous iteration's estimate by less than  $\epsilon$ . Each of these parameters, ( $\alpha$ ,  $n_{iters}$ , and  $\epsilon$ ) impact the GD solution in different ways.

In Fig. 3, the initial guess,  $x_0$ , is varied to show that a close initial guess leads to convergence in only a few steps. As the norm of difference between initial and final estimates grows, the number of iterations increases. For certain initial guesses, we observed that the algorithm did not converge in some

maximum number of iterations. For the negative Gaussian, this is because the gradient is very small far from the mean.

The convergence limit determines when the GD algorithm stops, seen in Fig. 2. The error increases if  $\epsilon$  is too large, as on the right of Fig. 1.

In GD, the norm of the gradient decreases as the solution converges, which is shown for the quadratic bowl function in Fig. 5.