Optimization using Gradient Descent

Consider the problem

$$\min_x f(x)$$

where,

$$f : R^d \longrightarrow R \quad (d > 0)$$

f is differentiable, and we are unable to analytically find a solution in closed form.

"Gradient descent is a first order optimization algorithm. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient of the function at current point."

$$f(x) = c \quad \text{(level surface/curves called contours)}$$

gradient of f (∇f) is orthogonal to the contours (normal to the surface)

Let us consider $f(x)$, a surface with ball starting at point x_0. When the ball is released, it will move downhill in the direction of steepest descent. Gradient descent exploits the fact that $f(x_0)$ decreases fastest if one moves from x_0 in the direction of the negative gradient, $-((\nabla f)(x_0))^T$ of f at x_0. Then, if

$$\underline{x}_1 = \underline{x}_0 - \alpha((\nabla f)(x_0))^T$$

for a small step size $\alpha \geqslant 0$, then $f(x_1) \leq f(x_0)$

This observation allows us to define a simple gradient descent algorithm:

If we want to find a local optimum $f(x_*)$ of a function $f: \mathbb{R}^n \to \mathbb{R}$, $x \mapsto f(x)$, we start with an initial guess $\underline{x}_0$ of the parameters we wish to optimize and then iterate as follows:

$$\underline{x}_{i+1} = \underline{x}_i - \alpha((\nabla f)(x_i))^T$$

for suitable α_i, the sequence $f(x_0) \geq f(x_1) \geq \cdots$ converges to local minimum.