Gradient methods in optimization

Gradient methods are popular techniques in numerical analysis and machine learning. Gradient descent / ascent are iterative algorithms that update parameters in the direction of the negative / positive gradient of a function to optimize it.

Consider a function $f(\theta)$ of some parameter or set of parameters θ. Gradient descent updates θ iteratively as follows:

$$\theta^{(t+1)} = \theta^{(t)} - \eta \nabla f(\theta^{(t)}),$$

$\nabla f(\theta)$ is the gradient and $\eta > 0$, a scalar, is the learning rate.

eg. minimizing a log-likelihood function:

$$\theta^{(t+1)} = \theta^{(t)} + \eta \nabla \log L(\theta^{(t)})$$

with $L(\theta)$, the likelihood function of interest.

Efficient for large data sets or high-dimensional optimization problems.

See: Deep Learning; I. Goodfellow, ...; 2016
Convex optimization; S. Boyd ...; 2004